Mechanistic and Statistical Models for Treatment and Control of Infectious Diseases

Mélanie Prague

Document présenté en vue d'obtenir le diplôme

Habilitation à Diriger des Recherches

École doctorale de Science Politique et Santé Publique de l'Université de Bordeaux
Spécialité Biostatistiques

par

Mélanie PRAGUE

MECHANISTIC AND STATISTICAL MODELS FOR TREATMENT AND CONTROL OF INFECTIOUS DISEASES

Soutenue publiquement le 5 Juin 2024, devant le jury suivant :

Samuel ALIZON	Directeur de recherche CNRS, Centre Interdisciplinaire de Recherche en Biologie CIRB, Collège de France, Paris, France	Rapporteur
Frederik GRAW	Professor, Modelling of Immune Processes, University Hospital Erlangen, Germany	Rapporteur
France MENTRÉ	Professeure des Université en Biostatistiques, Université Paris Cité, Hôpital Bichat, Inserm Infection Antimicrobials Modelling Evolution U1137, Paris, France	Rapportrice
Marie DOUMIC	Directrice de recherche Inria, Modelling and Analysis for Medical and Biological Applications, CMAP École polytechnique, CNRS, Palaiseau, France	Examinatrice
Cécile PROUST-LIMA	Directrice de recherche Inserm, Biostatistiques, Bordeaux Population Health U1219, Université de Bordeaux, France	Examinatrice
Joshua SCHIFFER	Professor, Vaccine and Infectious Disease and Clinical Research Divisions, Fred Hutchinson Center Seattle, USA; Department of Medicine, University of Washington, Seattle, USA	Examinateur
Rodolphe THIÉBAUT	Professeur des Universités - Praticien Hospitalier, Inria, Inserm Bordeaux Population health U1219, Statistics in systems biology and translationnal medicine, Université de Bordeaux, France	Garant

REMERCIEMENTS

Mes premiers remerciements s'adressent à **Rodolphe Thiébaut**, qui n'a cessé de m'encourager tout au long des ces années. Je lui suis reconnaissante pour sa confiance. Il a su me guider tout en me laissant une grande liberté dans mes choix et mes projets. Il est l'une des personnes au travail les plus bienveillantes que je connais et place toujours l'humain au premier plan. Je tiens également à remercier **Daniel Commenges** et **Victor de Gruttola**, qui ont successivement encadré mes travaux en doctorat et en postdoctorat. Leur passion pour la science et leur rigueur ont été pour moi une source d'inspiration constante. **Daniel**, votre curiosité scientifique n'a cessé de m'enrichir et de m'inspirer. Chaque jour, je mesure l'importance de vos enseignements.

Je tiens également à exprimer ma profonde gratitude à mes rapporteurs de HDR, **France Mentré**, **Samuel Alizon** et **Frederik Graw**, qui ont bien voulu examiner mes travaux et fournir des commentaires précieux. Leur intérêt pour ma recherche me touche profondément. Mes remerciements s'étendent aussi à **Marie Doumic**, **Cécile Proust-Lima**, **Joshua Schiffer** et **Rodolphe Thiébaut**, pour avoir accepté de siéger au jury de ma soutenance. Votre présence m'honore et, j'en suis convaincue, suscitera de nombreux échanges dans des domaines variés. Cela participera à me guider vers une recherche de plus en plus qualitative et impactante.

Je souhaite exprimer ma gratitude la plus sincère à l'égard de mes collaborateurs et collaboratrices, qu'ils soient mathématiciens, statisticiens, biologistes ou cliniciens. Le travail pluridisciplinaire occupe une place centrale dans ma démarche de recherche, et sans votre contribution, je ne pourrais envisager une recherche aussi complète. Je remercie tout particulièrement. **Jérémie Guedj**: pour ta vision pragmatique de la modélisation, ancrée dans une solide méthodologie. Collaborer avec toi dans l'encadrement d'étudiants est un plaisir, nos expertises se complètent à merveille. **Yves Levy**: pour tes retours toujours pertinents en immunologie, enrichissant nos recherches de questions biologiques cruciales et d'un regard critique sur les résultats et leur applicabilité. **Marc Lavielle**: pour les perspectives que tu as su m'ouvrir. Suivre ta pensée exige souvent une réflexion approfondie, mais travailler et discuter avec toi m'a montré de nouveaux horizons méthodologiques. **Alison Hill**: Despite a break due to our respective maternity leaves, our exchanges remain stimulating, and your passion for modeling encourages me to question further. Please come and visit again soon with your family. **Elisabeth Turner**: For our repeated reflections on the impact of missing data. Hope to be in town next time you come to Bordeaux. **Jane Heffernan**: Thank you for inviting me to your team meetings. Although I wish to make it more often, I really enjoy this opportunity. **Quentin Clairon**: d'abord postdoc puis collègue, pour m'avoir initié aux méthodes de contrôle optimal et être toujours prêt à démêler une démonstration complexe. **Xavier Hinault**: qui finira un jour par m'éclairer totalement sur la magie des réservoirs. **Laura Richert**, **Edouard Lhomme** et **Linda Witkop**: pour les discussions autour de projets appliqués et pour votre aide précieuse dans l'interprétation. Notre collaboration se renforce et j'espère qu'elle continuera de s'épanouir. **Boris Hejblum**: les bons conseils stats et pour toutes les discussions scientifiques et moins scientifiques. **Annabelle Collin**: pour m'avoir initiée aux méthodes d'assimilation de données avec beaucoup de bonne humeur. **Cécile Proust-Lima**: pour ta rigueur, ton engagement et tes méthodes de travail qui sont très enrichissants, je prends un réel plaisir à co-encadrer à tes côtés. **Virginie Rondeau**: partager le même bureau m'a permis de te poser mille et une questions avec des réponses toujours pertinentes.

Je souhaite exprimer ma gratitude envers l'ensemble de l'**équipe Sistm**. C'est un groupe unique, composé de nombreuses personnes talentueuses et toujours de bonne humeur. Un merci particulier à **Marta Avalos** et **Robin Genuer**, qui contribuent grandement à rendre notre environnement de travail agréable. J'espère que l'avenir nous permettra de mener des projets en collaboration plus

étroite. **Anton Ottavi**, j'espère que tu nous aideras à concrétiser tous nos rêves de projets. Mes remerciements s'adressent également à **Mélanie Huchon** pour son excellente gestion des activités extra-professionnelles, enrichissant ainsi notre quotidien. Finalement, mais l'un des plus grand merci à **Sandrine Darmigny** pour son engagement dans l'organisation de TOUTES nos activités côté INSERM, et pour sa bienveillance constante dans la gestion de mes demandes. Je tiens aussi à remercier **Audrey**, **Anne-Lise** puis **Ellie** pour leur succession efficace côté Inria. Interagir avec vous est toujours un plaisir.

Je veux adresser une reconnaissance toute particulière à ceux qui m'ont fait ou me font l'honneur de partager l'expérience d'une thèse ou d'un encadrement avec moi : **Patrick**, **Laura**, **Chloé**, **Marie**, **Auriane**, **Adrien**. Ce manuscrit vous doit beaucoup, et c'est pour moi un privilège d'avoir été ou d'être témoin de votre évolution. À **Marie Alexandre**, spécifiquement, merci pour ton engagement et ton sérieux dans toutes les missions qui t'ont été confiées. Ta fiabilité est indéniable, et j'ai eu beaucoup de chance de t'avoir comme première étudiante en co-direction de thèse. Merci aussi au différents étudiants qui se sont succédés en stage.

Pour conclure, je tiens à exprimer ma profonde gratitude envers l'ensemble des membres de l'**Inserm Bordeaux Population Health**, de l'**Inria Bordeaux** et du **Vaccine Research Institute**. Sans le soutien de ces institutions, il me serait difficile de trouver un tel épanouissement professionnel. Les interactions que j'entretiens avec mes collègues, que ce soit lors de nos séminaires conjoints ou autour d'un café, sont toujours extrêmement enrichissantes. Merci **Pierre Joly** pour les coucous du matin. Je pense tout particulièrement aussi aux équipes du SED (**Dan** et **François**), qui nous accompagnent souvent dans l'utilisation du centre de calcul Plafrim. Je remercie également les membres des équipes Inserm **Biostat**, ainsi que celles d'Inria **Monc**, **Mnemosyne** et **Flowers**. Ma gratitude envers vous est immense, même si vos prénoms ne sont pas tous mentionnés ici.

Je tiens à remercier chaleureusement mes amis, car la vie ne serait définitivement pas la même sans eux ! Un remerciement particulier à **Anthony**, qui, au-delà d'être un ami, est aussi un véritable guide spirituel. Grâce à lui et à mon engagement dans le patinage, j'ai pu enrichir mes connaissances dans ce sport, mais aussi dans les intrigues en tout genre et la gestion de l'humain. Merci à tous mes collègues **officiels d'arbitrage** pour les bons moments passés au bord des pistes. Je suis reconnaissante envers **Amandine** pour son soutien amical et émotionnel constant, ainsi que pour les nombreux moments partagés au téléphone. **Linda** merci pour tous ces moments partagé et l'amitié créé à Boston pendant notre séjour commun à Harvard. Nos retrouvailles familiales estivales sont des moments que j'espère voir se perpétuer d'année en année. À **JB**, **Harmonie**, **Sylvie**, et **Jean-Charles**, ma famille de cœur. **Vaea** et **Julia**, pour toutes les aventures qui nous attendent encore autour de la plage. **Loohan** et **Mélissa**, pour tous les voyages qu'il nous reste à faire. Un grand merci à **Benoît**, **Claire** et **Thomas** qui égayent notre quotidien et rendent nos week-ends festifs. **Charles** et **Annabelle** pour tout ce que nous partageons et j'espère que nous nous verrons souvent par la suite entre Bordeaux et Nantes, et ce plusieurs fois par ans. **Anaïs**, je te suis reconnaissante pour ces déjeuners où nous pouvons parler de tout. **Boris**, ta capacité à partager tant de rires au travail est précieuse. Notre amitié dure depuis des années, et est très importante pour moi. Finalement, j'aimerais remercier **Bruno** et **Diane** qui nous ont accueillis à Saint-Martin en décembre. Sans eux, l'écriture de cette HDR avec vue sur mer n'aurait pas été possible.

Je nourris une tendre pensée pour **ma famille**, qui, je le sais, se réjouit de mon bonheur. Je tiens à remercier mes parents pour leur soutien indéfectible et leur réconfort permanent. **Mon papa** qui a toujours été si fier de moi, même sans le dire, et **ma maman** qui a toujours été présente aimante et aidante pour m'accompagner dans les chemins que je souhaite prendre. Ils sont de véritables piliers de ma vie. Je vous aime. Ma gratitude s'étend également à ma **belle-famille**. Enfin, je garde une

pensée émue pour tous les **grands-parents** qui nous regardent de là-haut et qui, j'en suis sûre, sont extrêmement fiers.

Enfin, mes derniers et plus fervents remerciements s'adressent à **Vincent** et **Juliette** - mon époux et ma fille. Vincent, partager ma vie à tes côtés depuis plus de vingt ans représente un bonheur quotidien immense. Ce travail est également le tien, car tu me soutiens dans tous mes projets, professionnels comme personnels, sans jamais te plaindre des nombreux week-ends passés loin de toi dans les patinoires. Merci pour ton aide quotidienne inestimable. Merci pour l'amour que tu me portes - je suis si fière et comblée de t'avoir dans ma vie. Quant à Juliette, observer chaque jour ton épanouissement est une source de joie inépuisable. Tu es le rayon de soleil de mon existence. La vie se pare de couleurs toujours plus éclatantes au fur et à mesure que tu grandis. Même lorsque tu seras devenue une femme belle et intelligente, tu resteras toujours mon bébé. Je t'aime.

L'aventure de la HDR s'achève donc ici pour moi, avec ces derniers mots de remerciement rédigés après tous les autres. Dès demain, une nouvelle page se tournera, celle du reste de ma carrière de chercheuse. Mais avant de clore ce chapitre, je souhaite profiter de ces dernières lignes pour exprimer ma gratitude **à ceux que j'ai pu oublier de mentionner et qui n'en sont pas moins essentiels**.

Cette HDR commence ici pour les **lecteurs courageux** que je tiens, pour finir, à remercier.

Contents

Abbreviations

ABM: Agent-Based model
AIDS: Acquired Immunodeficiency Syndrome
ART: Antiretroviral Therapy
ATI: Analytical Treatment Interruption
BDF: Backward Differentiation Formula
CD4: CD4+ T cells
CHIMS: Controlled Human Infection Models
CoP/mCoP: Correlate of Protection / mechanistic Correlate of Protection
CRT: Cluster Randomized Trials
EBE: Empirical Bayes Estimates
GEE: Generalized Estimating Equations
HIV: Human Immunodeficiency Virus
IL-7: Interleukine 7
M(C)AR: Missing (Completely) At Random
MCMC: Monte Carlo Markov Chain
ML: Machine Learning
MSM: Marginal Structural Models
nAUC: time-averaged Area Under the Curve
NHP: Non-Human Primates
NLME: Non-Linear mixed Effects models
NPI: Non-Pharmaceutical Interventions
ODE: Ordinary Differential Equation
PLHIV: Person Living with HIV
RL: Reinforcement Learning
SAEM: Stochastic Approximation of the Expectation-Maximization algorithm
SBA: Sequential Bayesian Analysis
SCM: Stepwise Covariate Modeling
SIR: Susceptible - Infected - Recovered
VoC: Variant of Concern

Chapter 1

Experience in Research

1.1 Curriculum Vitae

1.1.1 Research Positions

MAR 2021- MAR 2023 | **Chaire Ecole Polytechnique, Sanofi**, Paris France.
Secondary appointment of research and teaching. Chaire "NUMERICAL INNOVATION AND DATA SCIENCE FOR HEALTH", work in collaboration with the Inria Team XPOP.
Reference: Marc lavielle

OCT 2016 - NOW | **Chargé de Recherche CRCN Inria** Bordeaux Sud-Ouest, France.
Researcher in the SISTM TEAM (Statistics In System biology and Translational Medicine). Also affiliated with Inserm U1219 Bordeaux Population Health.
Reference: Rodolphe Thiébaut

JAN 2014 - OCT 2016 | **Postdoctoral fellow Harvard school of Public Health**, Boston USA.
BIOSTATISTICS DEPARTMENT, working on "Semi-parametric estimation of treatment effect in clustered randomized trials in presence of missing data: application to HIV prevention in South Africa"
Reference: Victor de Gruttola

NOV 2013 - JAN 2014 | **Invited Researcher University of Oslo**, Norway.
INSTITUTE OF BASIC MEDICAL SCIENCES, working on "Comparison of propensity score based methods and dynamical methods for estimation of treatment effect in observational studies"
Reference: Odd Aalen; Jon Michael Gran

OCT 2010 - NOV 2013 | **Allocataire Moniteur Université de Bordeaux**, France.
TEACHING ASSISTANT position during my PhD (64 hours/year)
Reference: Pierre Joly

1.1.2 Diplomas

OCT 2010- NOV 2013	**PhD in Biostatistics**, Université de Bordeaux, France. "Monitoring of HIV infected patients based on dynamic models". * **With high honor from the jury (avec les félicitations).** * **Prix de thèse Société Française de Statistiques (2015)** Reference: Daniel Commenges
SEPT 2009 - SEPT 2010	**Master of Science Statistics and Econometrics**, Université de Rennes 1, France. at the department of mathematics, with high honors (dual curriculum)
SEPT 2007 - SEPT 2010	**Engineer Diploma ENSAI**, Rennes, France. National School for Statistics and Information Analysis.
SEPT 2008 - SEPT 2009	**Bachelor of Science in economics**, Université de Rennes 2, France. at the department of economics (dual curriculum)
SEPT 2005 - SEPT 2007	**Classes Préparatoires aux Grandes Ecoles** Lycée Montaigne, Bordeaux, France. MPSI/MP Math Physics and engineer sciences

1.1.3 Prizes and Awards

P-2024-1 TOP DOWNLOAD (RI-22-1) Paper in the 10% most downloaded among work published in an issue between 1 January 2022 – 31 December 2022 in CPT: Pharmacometrics & Systems pharmacology.

P-2021-1 **PEDR – Research "excellence" and doctoral supervision grant (2021 - 2025).**

P-2017-1 **PEDR – Research "excellence" and doctoral supervision grant (2017 - 2021).**

P-2016-2 CONFERENCE TRAVEL AWARD (SFDS) – This 600€-grant allow young scientists whose application get selected to attend a the French annual conference JdS (2016).

P-2016-1 HARVARD ROSE FELLOWSHIP AWARD – 4 prices a year to excellent scientific postdoctoral fellows to offer a cross-cultural experience of research in developing country. My project concerned: "Estimation of incidence and prevalence of HIV in Botswana: pooling data from different registers to evaluate the impact of personal characteristics such as citizenship".

P-2015-2 **THESIS AWARD - MARIE-JEANNE ET LAURENT DUHAMEL PRICE FROM SFDS is given once every other year and awards scientific quality of PhD works in Applied Statistics.**

P-2015-1 PHILIPPE FOUNDATION (RENEWAL) (6000$ personal grant) My project concerned: "Evaluating and targeting of HIV prevention strategies".

P-2014-1 PHILIPPE FOUNDATION Non-profit organization for Franco-American exchanges (5000$ personal grant). My project concerned: "Methods for analyzing HIV clustered randomized data and cohort data with informative missingness".

1.1.4 PhD Supervision and Mentoring (6)

- **Auriane Gabaut**, expected defense end 2026: Her PhD, tentatively titled "Methods for latent variable models in mechanistic models", is being conducted under the **co-supervision of Cécile**

Proust-Lima at Inserm Bordeaux Population health Biostat team and myself Inria/Inserm SISTM in Bordeaux. She is funded by an Inria CORDI-S grant and the PHDS network Impulsion (NG-22-1). The project endeavors to reconcile mechanistic modeling, traditionally reliant on limited markers, with the analysis of high-dimensional data, typically comprising extensive readouts from a few individuals. The initial phase involves enhancing the SAMBA algorithm to accommodate a high number of covariates in the statistical model of the mechanistic model, incorporating dimension reduction techniques, particularly lasso methods. Subsequently, the PhD work will focus on developing new statistical tools using latent variable models. This entails formulating a complex observation model for the mechanistic model, necessitating the creation of new likelihood estimators and further dimension reduction strategies. The ultimate goal is to develop tools applicable to mechanistic models, regardless of whether they possess an analytical solution. The application is anticipated to utilize data from the PREVAC-UP (IG-19-2) and Coverage-Immuno (IG-20-1) projects. Particularly in Coverage-Immuno, innovative methods have enabled the study of gene expression through sequencing from blood droplets collected via prick tests by patients themselves. This breakthrough facilitates daily sampling of gene expression, offering insights into the abundance of cell populations and the dynamics of biological metabolic pathways. These markers can be integrated into mechanistic models, enhancing the understanding of biological processes. She received the Best Poster award at the "2023 Bordeaux Population Health - Journée des jeunes chercheurs".

- **Adrien Mitard**, expected defense end 2026: His PhD, tentatively titled "Accelerating Vaccine Development Using Pre-Clinical Data", is being conducted under the **co-supervision of Jeremie Guedj at the Inserm IAME in Paris and myself Inria/Inserm SISTM in Bordeaux**. Adrien Mitard is located in Paris. He is funded by the PEPR Santé numérique SMATCH (NG-23-1). The endeavor is strategically positioned at the intersection of current SARS-CoV-2 challenges, with a dedicated focus on constructing predictive models. These models will be pivotal in assimilating data on viral load kinetics and immune responses using mechanistic models, thereby enhancing the strategic framework for vaccine development and the optimization of antiviral dosing regimens. The ultimate objective is to utilize these models to extrapolate from empirical data gathered in pre-clinical trials to human populations. The project encapsulates three principal initiatives: 1/ Joint Model Construction: The first initiative is to construct a comprehensive model that synergistically represents the dynamics of viral load and the humoral immune response following viral challenges in non-human primates. 2/ Innovative Pre-Clinical Study Designs: The second initiative seeks to propel the methodological framework of pre-clinical studies forward. This includes advocating for the implementation of lower innoculum levels and the strategic timing of challenges, informed by the expected dynamics of the investigated correlates of protection. 3/ Translational Research: The final initiative focuses on the translation of these pre-clinical findings to human contexts. This will involve leveraging allometric scaling principles and predictive modeling techniques to bridge the interspecies divide, thus enhancing the reliability of vaccine efficacy predictions for human populations.

- **Iris Ganser**, expected defense end 2024: Her PhD entitled "Effectiveness of non-pharmaceutical interventions and vaccines during the COVID-19 pandemic" is supervised by David Buckeridge from Mc Gills University and Rodolphe Thiébaut within the Inria/Inserm SISTM Team. It was funded by IdEX Bordeaux and the Fonds de Recherche Santé Québec. **I am a co-mentor**. The PhD projects are three-fold. 1/ Estimation of Non-pharmaceutical Interventions and Vaccines' Effects against COVID-19. This project aims to develop and utilize dynamical models to model the pandemics. I weekly supervised Iris on this topic and this was published in the Epidemics Journal (RI-24-1). 2/ Comparison of Regression and Mechanistic Models. The objective is to assess each model's effectiveness in capturing and predicting COVID-19 dynamics accurately. **I developed the idea of this second project.** 3/ Real-life Vaccine Effectiveness in Ontario,

Canada. I am only a consultant on this third project. The goal is to utilize data from blood donors and to perform mechanistic modeling of antibodies response to help in better understanding waning immunity. She received the Best Presentation award at the "2023 Bordeaux Population Health - Journée des jeunes chercheurs".

- **Marie Alexandre**, defended May 2022: Her PhD entitled “Mechanistic modeling and optimization of vaccine response in infectious diseases: Application to HIV, Ebola, SARS-CoV-2” was **co-supervized by myself and Rodolphe Thiébaut at Inria/Inserm SISTM team in Bordeaux**. She was funded by an Inria CORDI-S grant. Marie Alexandre focuses on the role of mathematical modeling in understanding the immune response induced by vaccination. It consisted un two major modeling pieces. First, the modeling of viral response in NHP after challenge with an infectious agent allows us to derive a method to define correlates of protection against infection. As far as methodology is concerned, we ground based our work into a causal framework (RI-22-4). Second, we modeled humoral response to vaccination in order to better understand and predict the long-term persistence of antibodies response after vaccination (RI-23-7). Finally, we developed a more biostatistical approach to analyse the data of vaccine trial for HIV using antiretroviral treatment interruption that are prone to informative right censoring (RI-21-2). This PhD led to 4 publications that I directly supervised and conceptualized (including RI-22-2). Marie Alexandre is now continuing as a Postdoctoral fellow in the Inria SISTM team.

- **Laura Villain**, defended Dec. 2018: Her PhD entitled "Analyse et modélisation de l’effet de l’Interleukine 7 (IL-7) chez les patients infectés par le VIH" was supervised by Daniel Commenges and Rodolphe Thiébaut within the Inria/Inserm SISTM Team. It was funded by the doctoral network of EHESP (Ecole des hautes études en Santé publique). **I was a co-mentor. Although not official at the doctoral school, I interacted weekly (and more) with Laura Villain during the second half of her PhD.** It builds on the seminal paper from Thiébaut et al. "Quantifying and Predicting the Effect of exogenous IL-7 on CD4+T cells in HIV-1 Infection." (RI-14-1) which launched a whole theme in the team on the modeling and optimization of delivery of IL-7 in individuals living with HIV. Interleukin-7 (IL-7) is a cytokine, which is a type of signaling molecule in the immune system. It plays a crucial role in the development and maintenance of T cells, a key component of the adaptive immune response. In the context of HIV infection, where there is a significant depletion of CD4+ T cells, IL-7 has been investigated as a therapeutic agent to help restore these cells. Its ability to enhance T cell survival and proliferation makes it a promising candidate for supplementing antiretroviral therapy in patients who do not achieve full immune reconstitution with antiretroviral therapy alone. The PhD of Laura Villain places in the context of the the INSPIRE trials series (phase I/IIa). It consisted in modeling the data to mechanistically understand and quantify the effect of IL-7 on survival and proliferation of T cells, as well as proposing optimal strategies to deliver cycles of IL-7 to HIV infected patients. In particular, we co-designed a Bayesian strategy to optimize the timing of IL-7 cycles in adaptive protocols (RI-18-2). She is now a scientist consultant in industry (NovaDiscovery then esqLABS GmbH).

- **Patrick Staples**, defended Fev. 2018: His PhD entitled "On the Statistical Properties of Epidemic Processes in Networks" was fully supervised by JP. Onnela at the Biostatistics department of Harvard School of Public health. During my postdoctoral fellowship, I had the opportunity to help supervise Patrick Staples for his second paper of his dissertation. **This consists in meeting twice a week (a one-to-one meeting and a meeting with the student’s mentor) for few months in order to conceptualize, perform the data curation and data analysis** of "Leveraging Contact Network Information in Clustered Randomized Studies of Contagion Processes" (RI-23-2). The basic idea was to understand which network features are the most

important to collect (number of contacts, status of contacts, ... etc) when evaluating a treatment effect against an epidemic that spreads on a network. He is now VP of Data Science at Alto Neuroscience USA.

1.1.5 Postdoc and Research Scientist Supervision (5)

- **Quentin Clairon** (Feb. 2019 to Oct. 2022): He was a postdoctoral fellow within the EBOVAC 3 Project (IG-18-1). His mission was to develop a new tool for estimation of parameters in non-linear mixed-effects models based on optimal control (RI-23-6). He also worked on the modeling of humoral response after vaccination (RI-23-5). Quentin Clairon is now a researcher ISFP at Inria within the SISTM team.

- **Marie Alexandre** (since May 2022): After her PhD, she is now a postdoctoral fellow. She continues to work on mechanistic modeling and dedicates a large amount of her time to the statistical analysis and design of pre-clinical (NHP and rodents) and clinical trials promoted by the Vaccine Research institute.

- **Myrtille Richard** (March 2022 to Dec. 2022): She was an engineer within the NIPAH Project (IG-18-2). To propose a pipeline for in silico trials, Myrtille explored multiple platforms for simulating mechanistic models. Specifically, we investigated the R packages deSolve and RxOde, the Python package diffeqpy, and Lixoft's Simulx software. This simulation platform was then used to generate data for designing adaptive trials and optimizing trial designs using the multi-armed bandit model [109]. Myrtille is now continuing her curriculum of Public Health Medical Doctor.

- **Maria Prieto** (Sept. 2020 to Sept. 2021): She was an engineer in the SISTM team. She worked on the modeling of the COVID19 epidemics in France and particularly helped in estimating the effect of non-pharmaceutical interventions of the epidemics dynamics. She is now a "Head of projects" at Octopia.

- **Dan Dutartre** (Part time May 2020 to Nov 2020): During the COVID-19 pandemics, Dan Dutartre has been partly detached by Inria to work within SISTM team. This was within the framework of the grant GESTEPID (NG-20-1). He designed a connector between R and Monolix that was particularly useful to share our codes during the pandemics. Since then, Lixoft has released the R "Lixoftconnectors" package that was basically providing similar functionalities. Dan is now on a long-term contract as research engineer at Inria.

1.1.6 Teaching Assigments

1.1.6.1 Teaching

I teach in average up to 50 hours of classes a year in statistics and advanced statistics at all levels from bachelor, masters to doctoral level. My main topics for teaching are: basic statistics, dynamical/mechanistic models and missing data.

- Master level
  - **ENSAI (2016 - now):** Full coordination of module on "missing data" for 3rd year students, option Biostatistics (16h).
  - **Master 2 Biostatistics University Bordeaux (2016 - now):** Module on "Multiple imputation" (8h)

  - **Master 2 Polytechnique MSV Modélisation Science du vivant (2021 - 2023):** Full coordination of module on "Mixed-effects models for population approaches" for Master 2 students (28h).
  - **Master 2 Biostatistics University Bordeaux (2019 - 2022):** Module on "Introduction to dynamical models for epidemics" (6h).
  - **ENSAI (2016 - 2018):** Full coordination of module on "Epidemiology and article critical reading" for 3rd year students, option Biostatistiques (14h).
  - **ENSAI (2017 - 2020):** Module on "Introduction to Research" for 3rd year students, option Biostatistiques (5h).
  - **Master 2 Pharmacology, University Bordeaux (2017 - 2019):** Design experiment (18h)
  - **Master 1 Stochastic Processes, University Bordeaux (2017 - 2019):** Bayesian Stat (7h)

- University Diploma & Summer School
  - **DU Méthodes statistiques de régression en épidémiologie (2017 - now):** Module on "Linear regression and analysis of variance". Correspondance continuous education (equiv. 12h).
  - **ISPED Summer school (2017- now):** Initiation to R. Continuous education. (between 4h and 14h depending on years)
  - **PhD module Modeling life science (2018):** Methods for ODE/PDE/SDE, organization & coordination of all lecturers 35h – lecture 7h

1.1.6.2 Internships (18)

I supervised a total of 18 intern students, among them one bachelor level, 6 master 1 level, 10 master 2 level and 1 PhD level.

- **2023 (4)**
  - Marie COLIN (ENSAI 2 ieme année June – Aug. 2023) "Mechanistic modeling of antibodies response using transcriptomics data" . This work lead to one international conference talk CI-23-2.
  - Junior JUMBONG (ENSAI 2 ieme année June – Aug. 2023) "Analysis of determinants of dynamical vaccination response using mechanistic models".
  - Fanny MOREAU (Licence 3 MIASCHG Bordeaux May – July 2023) "Analysis of determinants of vaccination response at 12 months using regression models".
  - Auriane GABAUT (Master 2 Modélisation science du vivant Polytechnique Paris Apr. – Sept. 2023) "Model building in nonlinear models in presence of a high number of covariates". This work lead to one national conference talk CN-23-1.

- **2022 (2)**
  - Marie POUPELIN (Master 1 Biostatistiques Université bordeaux, Apr. – July 2022) : « Humoral response to ebola vaccination in prevac-Up trial ». This work helped with the data management of one submitted paper Sub-3.
  - Florian ROBERT (Master 2 Paris Sorbonne Data science June - Oct 2022) : « Model building strategies in NLME »

- **2021 (3)**

  - Marie-Laure CHARPIGNON (PhD Candidate MIT June – Oct 2021) : "Evaluation of the impact of non-pharmaceutical interventions on SARS-CoV-2 transmission rate at the department level in France". This work led to the publication RI-23-3.
  - Carole VIGNALS (Master 2 epidémiologie Université bordeaux, Interne Santé publique Apr. – Oct 2021) : « Relâchement des gestes barrière en France à l'ère de la vaccination contre la COVID-19 ». This work led to the publication RI-21-4 and public outreach RN-23-1.
  - Abdelghani NEHAUS (Master 2 Bioinformatique Labri Université bordeaux Apr. – Aug. 2021) : "Optimization of intervention strategies in epidemic models using deep reinforcement learning techniques"

- **2019 (3)**
  - Guillaume SOTTON (Master 2 Bioinformatique Labri Université bordeaux Apr. – Aug. 2019) : " Joint fragility model based on differential equations: analysis of marker kinetics and survival time "
  - Clément LEMOIGNE (Mathmeca 2nd year July - Sept. 2019) : "Comparaison de méthodes de construction de modèles et de sélection de variables pour les modèles dynamiques en santé. "
  - Marie ALEXANDRE (Master 2 Centrale Marseille, Apr. – Aug. 2018) : "PKPD modeling in pre-clinical development of T cell bispecific immunotherapy". This was a co-supervision with Industry (Nicolas Frances, Roche, Basel). Marie continued as a PhD Student co-supervized by myself.

- **Before 2018 (6)**
  - 2018 - Marie ALEXANDRE (Master 1 Centrale Marseille, Jun. – Aug. 2017) : "Modèle Vaccinal pour Ebola Dynamique des anticorps et de la réponse cellulaire".
  - 2017 - Paul TAUZIA (Master 2 MSS Université de Bordeaux, Apr. – Sept. 2017): "Utilisation de la déconvolution cellulaire pour détecter des différences d'expression génique".
  - 2017 - Augusta ALPHONSE (Master 2 in Vaccinology Creteil VRI, Apr. – Sept. 2017) : "Proof of concept for an automated gating tool applied to flow cytometry data from a HIV therapeutic vaccine trial"
  - 2013 - Damien FOSSAT-CERCLER (Master 2 in Statistics and ENSAI 3rd year, Apr. – Sept. 2013) : "Dynamical modeling of pharmacoepidemiology in observational studies: optimization of methods for numerical integration"
  - 2012 - Ana JARNE (Master 2 in Biostatistics, Université de Pau, Apr. – Sept. 2012) : "Modeling of T-lymphocytes dynamics after InterLeukine-7 injections in HIV infected patients: introduction of a feedback loop". This work was continued as a PhD and led to the publication RI-17-3.
  - 2011 - Sybille MASSE (ENSAI 2nd year, Jun. – Sept. 2011) : "Understanding the Effect of Interleukin-7 with mathematical models"

1.1.7 Expertizing activities

- **Editorial Board:** I am Associate Editor for the Journal "**International Journal of Biostatistics**".

- **International Expert:**

  - MILLENNIUM SCIENCE INITIATIVE (Chile) Application reviewer for the Natural/Exact Sciences application (9 projets) **(2017; 2019; 2021; 2023)**.
  - ANR-DFG GERMAN RESEARCH FOUNDATION (France-Germany) Application reviewer (2 projets) **(2018)**.
  - SWISS NATIONAL SCIENCE FOUNDATION (SNSF) (Suisse) Application reviewer (8 projets) **(2020; 2021; 2022; 2023)**.

- **National Expert:**
  - AGENCE NATIONALE DE LA RECHERCHE (ANR) CES45 Member of evaluation commitee **(2023 - now)**
  - AGENCE NATIONALE DE LA RECHERCHE SUR LE SIDA ET HÉPATITES VIRALES ET LES MALADIES ÉMERGENTES (ANRS MIE) CSS13 "Clinical research" Member of evaluation commitee **(2019 - now)**.
  - SOCIETÉ FRANCAISE DE STATISTIQUES (SFdS) Member of the Jury Award Marie-Jeanne Laurent-Duhamel **(2022 - now)**.
  - AGENCE NATIONALE DE LA RECHERCHE SUR LE SIDA ET HÉPATITES VIRALES (ANRS) CSS3 "Clinical research on HIV and comorbidities" Member of evaluation commitee **(2016 - 2019)**.
  - ANR COVID RA Covid-19 édition 2021 Vague 14 (2 projects) **(2021)**.

- **Member of International Scientific Committee:**
  - Member of Scientific Director board of the COVID IMMUNITY TASK FORCE IN CANADA **(2021 - 2023)**.
  - 4th WORKSHOP ON VIRAL DYNAMICS Paris **(Oct. 21-23 2019)**.

- **Member of National Scientific Committee:**
  - JDS French statistical society annual conference Nancy **(3-7 June 2019)**.
  - Conference STATISTICS IN HEALTH – personalized medicine, CIMI, Toulouse **(Jan. 11-12 2018)**.

- **Reviewer:**
  - JOURNALS IN STATISTICS AND BIOSTATISTICS: Biometrics (since 2014); JRSS-B (since 2016); Statistical Methods in Medical research (since 2017); JRSSA Interaction (since 2019); Journal of royal society open science (since 2021)
  - JOURNALS IN COMPUTATIONAL BIOLOGY: Plos One (since 2014); IEEE/ ACM Computational biology and BioInformatics (since 2010); elife (since 2021); CPT PsP (since 2022); plos Computational Biology (since 2023)
  - JOURNALS IN APPLIED MATHEMATICS TO MEDICINE: Statistics in medicine (2013); Trials (since 2016); Society of clinical trials (2016); Statistical Science (since 2019)
  - See Web of science Peer review (since 2021 only)

- **Hiring committees (6):**
  - MCF CNAM Section 26, Mai 2017, Paris
  - CRCN INRIA Bordeaux 2019, 2020 and 2024; Paris-Saclay 2022; Nice 2023

- **International PhD Defense as "Rapporteur" (1):**

  - STEVEN SANCHE, Reviewer, University of Montreal Faculty of Pharmacy. "Effet des antirétroviraux sur la pathogénése du VIH : une étude par modélisation mathématique intégrant la cinétique du virus, de l'immunité, du médicament, et le comportement d'adhésion avec leurs variabilités interindividuelles", Dec. 2018.

- **National PhD Defense as "Jury Member" (5):**
  - JOHANN FOUAZI, Examiner, Brain and spine institute, Inria. "Machine learning to predict impulse control disorders in Parkinson's disease." Dec. 2019.
  - ANTONIO GONCALVES, Examiner, Paris Diderot, IAME inserm, « Modélisation de l'effet de nouvelles molécules anti-VHB chez la souris et chez l'homme. » Feb. 2021.
  - ROMAIN NARCI, Examiner, Université Paris Saclay Inrae, « Inférence dans des modèles à effets mixtes pour des dynamiques épidémiques partiellement observées récurrentes et multisites » March 2022.
  - GUILLAUME LINGAS, Examiner, IAME Inserm, « Modélisation de la dynamique virale du SARS-CoV-2 : Implication pour l'évaluation thérapeutique » Oct. 2022.
  - THOMAS BENETTEAU, Examiner, IRD Montpellier, « Modélisation mathématique des infections HPV : quel rôle du hasard dans la persistance et l'oncogénèse ? » Dec. 2023.

- **PhD follow-up committees (10)** (once a year up to 3 years):
  - Chiara Nicolo (2017 - 2020; director Sébastien Benzekry - U. Bordeaux). Mathematical modeling of neoadjuvant antiangiogenic therapy and prediction of post-surgical metastatic relapse in breast cancer patients.
  - Jonas Beale (2018 -2019; director Aurélien Latouche and Emmanuel Barillot - University Pierre et Marie Curie Paris). De la modélisation mécanistique des voies de signalisation dans le cancer à l'interprétation des modèles et de leurs apports : applications cliniques et évaluation statistique.
  - Marie-Astrid Metten (2018 - 2020; director Guillaume Chauvet and Jean-Francois Viel - U. Rennes). Données manquantes dans les études de cohorte.
  - Imke Mayer (2018 - 2021; director Julie Josse and Jean-pierre Nadal - Ecole Polytechnique). Analyse de données hétérogènes avec données manquantes - Application à la prise en charge des polytraumatisés graves.
  - Iris Granger (2020 - 2024; director David Buckeridge and Rodolphe Thiébaut - Mc Gills Canada U. Bordeaux). Modeling the epidemics of COVID-19.
  - Baptiste Elie (2021 - 2024; director Samuel Alizon and Nacho Bravo - U. Montpelier). Etude des infection génitales HPV.
  - Benjamin Glemain (2022 - 2025; director Fabrice Carrat and Nathanaël Lapidus; Inserm iPLesp Paris). Corrélats de protection contre les différents variants du SARS-CoV-2.
  - Maxime Beaulieu (2023 - 2026; director Jérémie Guedj; Inserm Iame Paris). Modélisation de l'efficacité des stratégies antivirales contre les variants du Sars-CoV-2 : de la population générale aux patients hospitalisés.
  - Erwan Gaymard (2023 - 2024; director Maxime Sermesant and Irene Balelli - Inria Nice, Exactcure CIFRE). Développement de méthodologies mathématiques en méta-modélisation PK à partir de sources hautement hétérogènes.
  - Eve Rahbe (2023 - 2024; director Lulla Opatowski and Philippe Glaser - Institut Pasteur). Modélisation spatio-temporelle de l'antibiorésistance à l'échelle mondiale.
  - Ilona Suhanda (2023 - 2026; director Raphaëlle Metras - Sorbonne Université). Spatial risk assessment of lyme borreliosis and thick-borne enceohalitis: a joint modeling approach.

1.1.8 Administrative activities

- **Research coordination at a National level**
  - 2021 - now - MEMBER OF THE BOARD "ACTION COORDONNÉE MODÉLISATION" (ANRS-MIE).
  - 2021 - now Co-leader with J. Guedj of the working group WITHIN-HOST MODELING Action Coordonnée Modélisation (ANRS-MIE).
  - 2018 - 2021 - President of COMMUNICATION GROUP AT FRENCH STATISTICAL SOCIETY (SFDS).
  - 2016 - 2018 - Member of COMMUNICATION GROUP AT FRENCH STATISTICAL SOCIETY (SFDS).
  - 2016 - 2020 - Elected member of the Board of YOUNG FRENCH STATISTICIANS at French statistical Society (SFdS).
- **Local duties**
  - COMMISSION INRIA DE DÉVELOPPEMENT TECHNOLOGIQUE Evaluation and expertise of grants applications to Inria for the transfer of technology, partnership with industries and development of new technological tools (2016 - 2022).
  - COMMISSION INRIA EMPLOI-RECHERCHE Participation in the scientific committee for attribution of research grants for delegations, PhD and Postdoctoral fellowships (2017 - 2023).
- **Organisation of International Event/Conference**
  - 7TH WORKSHOP ON VIRAL DYNAMICS about 100 attendees Head of organizing and scientific committee 2025 in Bordeaux
  - 4TH WORKSHOP ON VIRAL DYNAMICS about 100 attendees in collaboration with Jeremy Guedj. Creation of the conference website, communication, monitoring of registration, conception of a booklet, Member of scientific committee. Oct 2019.
- **Organisation of National Event/Conference**
  - JOURNÉE DE LA STATISTIQUE 2024 Member of the communication group for the organization committee (about 500 participants) in Bordeaux May 2024
  - 2ND WORKSHOP AC MODÉLISATION ANRS MIE about 80 attendees Head of organizing committee 21-22 Nov. 2022 in Bordeaux
  - WORKSHOP ON WITHIN-HOST DYNAMICS about 50 attendees (30% international). Guest speaker Alan Perelson. 17 Sept. 2022 in Paris.
  - SCIENTIFIC LUNCHES AT JDS I created this type of event and organized it the three first years. It consists in organizing a lunch around a specific scientific topic as a satellite event of the French statistical Society conference (pick the location, invite people, stimulate discussion – around 20 attendees). This type of satellite event still exists at the conference. I personally organized those in 2017, 2018 and 2019.

1.2 Grant and Funding

1.2.1 Opening Remarks

In this funding section, the grants are labelled depending on the type, the year and the number of grants during the same year. The nomenclature writes "type-year-number". Types of grants are as follow:

- IG : International Grants
- NG : National Grants

1.2.2 International grants (5)

IG-20-1 **Coverage-Immuno** – EIT Health COVID-19 Rapid Response - COVERAGE-Immuno aimed at performing a deep, repeated evaluation of immunological markers and transcriptomics data in COVID-19 positive patients treated at home in the context of COVERAGE, a randomized clinical trial evaluating several experimental treatments at home (legal trial sponsor: Bordeaux University Hospital). The project will allow a better stratification of patients based on immuno biomarkers profiles so as to adapt their management and care before hospitalization.
Head: Rodolphe Thiébaut (Inserm)
Role & Specific tasks : I am workpackage leader for "Data analysis". We use the data to benchmark if it is possible to fit immune dynamics to SARS-CoV-2 response using transcriptomic data.
Dates: 04/2020 - 12/2020
Main partners: Inserm, Inria
Total COVERAGE budget: 471k€ **SISTM COVERAGE-immuno budget:** 89k€ **WP budget:** 13k€
Direct Personal outputs to date: CI-23-2; CN*-23-1

IG-19-2 **Prevac-Up**: European project RIA EDCTP - PREVAC-UP is built around the PREVAC consortium and its main objective is to evaluate the long-term safety, as well as the durability of humoral and cellular immune responses, of three different Ebola vaccine regimens previously tested by the consortium for 5 years after vaccination. PREVAC-UP also assesses the impact of co-infection on the immune response to vaccination, build on the extensive community mobilisation efforts previously generated through PREVAC to provide a transnational platform for social and health science research and training. It thus aims to expand and sustain capacity building and training of scientists in the four participant African countries.
Head: Yazdan Yazdanpanah (Inserm)
Role & Specific tasks : I am workpackage leader for "System Vaccinology approach". SISTM contributes to the integrative statistical analysis of the immune response which will be used to explore the mechanism of action of the vaccines and to identify early correlates of durable antibody induction.
Dates: 01/2019 - 06/2024
Main partners: Coordinated by Inserm (France). Other beneficiaries: CNFRSR (Guinea), CERFIG (Guinea), LSHTM (UK), COMAHS (Sierra-Leone), NIAID (USA), NPHIL (Liberia), USTTB (Mali), Centre pour le Développement des Vaccins (Mali), Inserm Transfert SA (France)
Total PREVAC-UP budget: 16M€ **SISTM budget:** 328k€
Direct Personal output to date: Sub-3

IG-19-1 **amfAR**: American Fundation for AIDS research; Impact Grant 109856-65-RGRL - Mechanistic and empirical modeling of viral rebound to identify predictors of post-treatment control.
Head: R. Wang (Harvard University)
Role & Specific tasks : I work as a contributor for mechanistic models aspects.
Dates: 01/2019 - 01/2020
Main partners: Harvard University, Inserm
Total amfAR budget: 110k€ **SISTM budget:** 7k€
Direct Personal output to date: RI-20-3

IG-18-2 **Nipah Virus Study**: Sino-French Agreement Aviesan - To raise the challenge caused by Nipah virus we propose to develop a program that shall led to a better understanding of the epidemiology of the virus as well as the associated physiopathology, to develop new tools in the field of diagnosis, treatment and prevention of the infection.
Head: Hervé Raoul (Inserm)
Role & Specific tasks : I am co-workpackage leader for "Biostatistics and evolutionary analysis". I am in charge of the modeling of vaccine response aspects in this project. SISTM contributes to model and analyse data from vaccine trial against Nipah virus in non-human primates. Because almost no data were generated because of COVID-19 pandemics, the work was retargeted to be more theoretical on model building strategies development.
Dates: 01/2018 - 12/2023
Main partners: Inserm, Université de Marseille, VRI, Wuhan institute of virology, Chinese center for disease control.
Total NIPAH budget: 1M€ **SISTM budget:** 150k€
Direct Personal outputs to date: RI-22-1; HCI-22-3; HCI-21-2

IG-18-1 **EBOVAC3** – European IMI2: The EBOVAC3 project aims to support an essential part of the remaining clinical and manufacturing activities required for the licensure of the two-dose vaccine regimen in the European Union and the United States. Building on work carried out under the EBOVAC1 and EBOVAC2 projects, EBOVAC3 is running clinical trials to gather safety and immunogenicity data in infants in Sierra Leone and Guinea; as well as health care workers in an area affected by Ebola in the Democratic Republic of Congo (DRC). It also follows up participants who received the vaccine regimen in the EBOVAC-Salone trial in Sierra Leone to assess the safety and immunogenicity of the regimen in the longer term. In addition to the clinical trials, EBOVAC3 aims to characterize outbreak preparedness in Sierra Leone, Guinea and the DRC through social science research and mapping of the existing preparedness activities in each of the three countries.
Head: Deborah Watson-Jones (LSHTM)
Role & Specific tasks : I am co-workpackage leader for "Modeling" aspects in this project. SISTM will contribute to modeling the long-term humoral response to the vaccine.
Dates: 06/2019 - 05/2023
Main partners: Coordinated by the London School of Hygiene and Tropical Medicine (United Kingdom). Other beneficiaries: Janssen a Pharmaceutical Companies of Johnson and Johnson, Inserm (France), The University of Antwerpen (Belgium), University of Sierra Leone (Sierra Leone)
Total EBOVAC3 budget: 51M€ (including 29M€ from EU) **SISTM budget:** 351k€
Direct Personal outputs to date: RI-23-6; HCI*-19-1; CI-19-4; RI-23-7; HCI-21-4.

1.2.3 National grants (6)

NG-23-1 **PEPR Santé Numérique SMATCH** – Addressing new methodological challenges in public health data science. The project aims to develop and apply statistical methods and AI-based approaches to accelerate the development of medical interventions in clinical trials, focusing on the early phases and integrating multi-source data.
Head: Sarah Zohar (Inria Heka) and Rodolphe Thiébaut (Inria/Inserm SISTM)
Role & Specific tasks : I am involved as a contributor and mentor of one PhD student.
Dates: 09/2023 - 08/2029
Main partners: Inria (Heka, Sistm, Premedical, Soda), Inserm, CNRS, CEA, CHU Bordeaux Tours Poitiers Montpellier, Université Paris Cité, Bordeaux, poitiers, montpellier, Tours, Lille, HAS

Total SMATCH budget: 1.4M€ **SISTM budget:** 350k€ **Own budget:** 130k€
Direct Personal output to date: None to date.

NG-22-1 **PHDS Impulsion** Réseaux Public Health Data Science network in Bordeaux University. The PHDS network aims to utilize Bordeaux's expertise in statistics, applied mathematics, and public health to develop analytical methods that: precisely predict disease onset and progression at an individual level, elucidate the mechanisms that underlie diseases, and create and assess innovative therapeutic strategies..
Head: Cécile Proust-Lima (Inserm)
Role & Specific tasks : I am the workpackage leader of the tack "Incorporate complex measurement models into theoretical mechanistic models".
Dates: 03/2022 - 03/2026
Main partners: Inserm BPH, Inria Bordeaux, LaBRI, IMB
Total Impulsion budget: 855k€
Direct Personal outputs to date: The WP produced [65] and CN-23-1 to date.

NG-21-1 **EMERGEN - Modvar:** Consortium for Surveillance and Research on EMERgent Pathogens via Microbial GENomics. EMERGEN, coordinated by Sante publique France and ANRS-Emerging Infectious Diseases, aims to deploy a genomic surveillance system for SARS-CoV-2 infections throughout France. Its main objective is to follow the genetic evolution of the SARS-CoV-2 virus in order to detect the emergence and the spatio-temporal distribution of variants, i.e. viruses with mutations likely to have functional consequences, such as infectivity, contagiousness, virulence or immune escape.
Head: Simon Cauchemez and Vittoria Collizza. Pasteur Institute
Role & Specific tasks : I am co-workpackage leader regarding the within-host aspects of the project. SISTM will contribute to modeling the impact on epidemic dynamics of SARS-CoV-2 variants based on the estimation of their within-host characteristics.
Dates: 12/2021 - 06/2024
Main partners: Santé Publique France, Inserm/ANRS, APHP, HCL, Pasteur Institute, Anses, IFB, CNRGH/CEA, Réseau Sentinelles
Total Emergen budget: 10M€ **Modvar budget:** 450k€ **SISTM budget:** 56k€
Direct Personal outputs to date: RI-23-8; RI-23-5

NG-20-1 **GESTEPID** : Action Inria stop COVID19 - Modeling of the COVID-19 epidemics in France with a focus on the region Nouvelle-Aquitaine.
Head: Mélanie Prague (Inria/ Inserm)
Role & Specific tasks : I am the principal investigator
Dates: 03/2020 - 12/2020
Main partners: Inserm, Inria, Santé Publique France
Total/SISTM budget: 6 months of engineer
Direct Personal outputs to date: RI-23-1; HCI*-21-1; HCI-21-3;HCI*-21-5; S-20-1; RI-24-1; HCI-23-5; CN-22-1; RI-23-3. This work was cited by the "Conseil Scientifique COVID-19" on his report June 2nd and gave birth to multiple press releases see Section 1.3.13.

NG-18-1 **DYNAMHIC** (Dynamical modeling of HIV Cure Duration): Inria associate team. This collaboration allows the analysis of unique pre-clinical data in non-human primates of HIV cure interventions.
Head: Mélanie Prague (Inria/ Inserm) Alison Hill (Harvard)
Role & Specific tasks: I am the principal investigator on the French side. This grant funded essentially travel to Boston and to Bordeaux as well as to conferences. I was in charge of the methodological aspects for inference using mechanistic models.

Dates: 01/2018 - 01/2022
Main partners: Johns Hopkins University School of Medicine, Harvard University
Total/SISTM budget: 10k€/year for travel expenses
Direct Personal outputs to date: RI-20-3; Sub-2; HCI-20-1; HCI-18-1; CI-19-5.

NG-17-1 **EVALUATES:** Funded by National Institute for Cancer. Development of joint models involving dynamical mechanistic model of tumor growth in colorectal cancer.
Head: Virginie Rondeau (Inserm Bordeaux Population Health)
Role & Specific tasks: I am work package leader on mechanistic modeling. This grant aims at funding 6 months of intern students, travel and equipment expenses in order to investigate the use of Monolix for fitting joint models and to compare it with an existing R package called "frailtypack".
Dates: 08/2017 - 06/2021
Main partners: Université de Bordeaux, Inserm, Inria, CHU besançon
Total Emergen budget: 192k€ **SISTM budget:** 6k€
Direct Personal output to date: Internship Guillaume SOTTON.

1.3 Scientific production

1.3.1 Spirit of my research

In next section, we offer a detailed overview of the scientific achievements accomplished over the past year. We also provide bibliometrics information in term of scientific production. Each research focus is concisely summarized, complemented by a table that encapsulates the completed work. Following this, we provide an in-depth exploration of how these contributions (referred by RI-XX-X; RN-XX-X; HCI-XX-X; CI-XX-X; CN-XX-X.. etc in Section 1.3) are placed in the existing literature and interrelate.

In the SISTM team, where I currently serve as a researcher, we conceptualize applied research in medical science and vaccine development as a cyclic process, akin to a wheel (refer to Figure 1.1). This is made possible by an unique partnership with the Vaccine Research institute. This process initiates with the design and execution of a trial, followed by the necessity of data management and statistical analysis. These steps pave the way for data integration, mechanistic modeling, and in silico predictions, which are crucial for planning subsequent trials. My research is primarily anchored in mechanistic modeling (for which I lead the scientific axis in the team) and the course of our research is largely contingent on the nature of the data obtained and the unique scientific challenges they present. Our focus has been predominantly on infectious diseases, with a more recent shift towards vaccine development. These topics are described in the section 2.1 for methodology and the section 3.1 for data analysis and development of advanced computational biology tools. Finally, an additional facet of my research, that has been inspired by my postdoctoral training at Harvard School of Public Health, extends beyond clinical trials. It aims to understand the possibilities that offer observational studies to better understand the effect of an intervention. This includes evaluating intervention and vaccine strategies within populations. This last topic is described in the section 4.1.

Together, these three axes provide a holistic and multi-faceted approach to advancing the field of vaccine development, offering both theoretical insights and practical applications that are vital for addressing current and future challenges. Since commencing my research career in 2010, I have contributed to the scientific community, as evidenced by my authorship or co-authorship of 40 articles. This includes 8 articles where I served as the first author, 6 as the second author, and 11 in the capacity of the last or before last author. Notably, all these publications are centered on the modeling

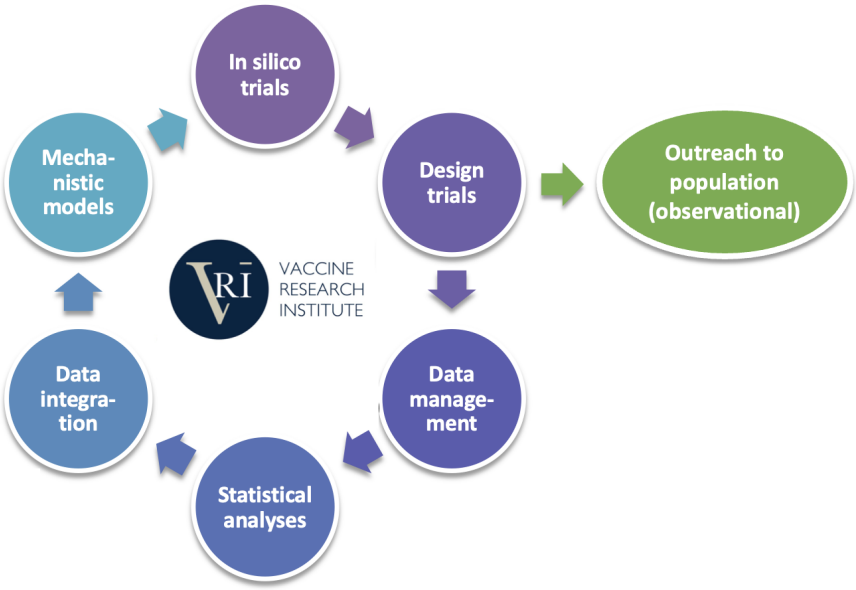


Figure 1.1 – Conceptual Framework of Applied Research in Data science for Medical research in the SISTM Team: Navigating from Trial Development to Analysis, Simulations in silico, Design of subsequent trials and outreach to population in an iterative loop.

of infectious diseases: 21 were published in biostatistical or applied mathematics/statistics journals and present methods development, while 19 appeared in journals focusing on computational biology or medicine.

1.3.2 Opening remarks

In this "own" bibliography section, the scientific production is labelled depending on the type, the year and the number of production during the same year. The nomenclature writes "type-year-number". Types of production are as follow:

- RI : Peer-reviewed international journals
- RN : Peer-reviewed national journals
- Sub : Submitted papers with available preprints
- HCI*/HCI : Invited or competitive international conferences
- CI : Other international conferences
- CN*/CN : Invited or not National conferences.
- S : Seminars in institutions other than my own affiliation.
- Soft : Software production.

Links with DOI are available to access the published articles in internationnal journals (RI - Section 1.3.3) and national journals (RN - Section 1.3.4), as well as all the preprints to justify their submission status (Sub - Section 1.3.5). In biostatistics, journal articles are more important than conferences. Order of authors depends on the contribution. Generally, the two first and the two last authors are the main contributors. For senior authorships (two last authors), medical doctors are usually placed last, after the biostatisticians. I tried to highlight my contribution for each journal

article.

In biostatistics again, there is barely never proceedings for conferences (other than 1000-words abstracts booklets). I separated the conferences in 5 categories: Invited in international conference (HCI* - Section 1.3.6), competitive international conferences for which there is a single-track and around 20% to 30% of acceptance (HCI - Section 1.3.7), regular international conferences (CI - Section 1.3.8), invited national conference (CN* - Section 1.3.9) and regular national conferences (CN - Section 1.3.10). In each case, I specified if I was the speaker or if it was one of my mentee or a collaborator. In general, I try to valorize each journal article in one or two conferences (either applied or more methodological) in order to network. Finally, I also listed seminars I gave in institutions other than my own affiliation (S - Section 1.3.11). In general, I try to give at least one seminar a year in my own lab meetings.

When applicable, I try to release software or open source code so that people can broadly use our methods (Soft - Section 1.3.12). In each article, the code for dissemination and replication is always available in supplementary material.

I would like to drive the attention of the fact that :

- Articles RI-23-2, RI-20-3, RI-19-2, RI-18-1, RI-17-5, RI-17-2, RI-17-1, RI-16-5, RI-16-2, Sub-2 are written with only international collaborators (Harvard and Duke university) without any team member of the SISTM team.
- Articles RI-22-5, RI-22-1, RI-20-2 and RI-23-8 are written with only national collaborators (Inria MONC team, M3DISIM team, XPOP team and Inserm IAME Team BIPID) without any team member of the SISTM team other than my mentee.
- Other articles are created within SISTM team, having other members as co-authors.

1.3.3 Peer-reviewed International journals (40)

RI-24-1 Estimating the population effectiveness of interventions against COVID-19 in France: a modelling study
Ganser I., Buckridge D., Hefferman J., **Prague M.** and Thiébaut R.
Epidemics (IF=5.3) - In press - Jan 2024 .
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Software Development; Writing; Proofreading; Mentoring.

RI-23-8 Impact of variants of concern on SARS-CoV-2 viral dynamics in non-human primates,
Marc A., Marlin R., Donati F., **Prague M.**, Keroui M., ..., Le Grand R. and Guedj J. **PLOS Computational Biology** (IF=4.8) - 19(8) - e1010721 - Aug 2023 (16 pages).
Contribution: Development of Methodology; Proofreading; Mentoring.

RI-23-7 Prediction of long-term humoral response induced by the two-dose heterologous Ad26.ZEBOV, MVA-BN-Filo vaccine against Ebola
Alexandre M., **Prague M.**, McLean C., Bockstal V., Douoguih M. and Thiébaut R. **NPJ Vaccines** (IF=9.2) - 8 - 174 - Nov 2023 (32 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Writing; Proofreading; Mentoring.

RI-23-6 Parameter estimation in nonlinear mixed effect models based on ordinary differential equations: an optimal control approach.

Clairon Q., Pasin C. Balelli I., Thiébaut R. and **Prague M.** **Computational Statistics** (IF=1.3) - in press - Oct 2023 (32 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Writing; Proofreading; Mentoring.

RI-23-5 Modeling the kinetics of the neutralizing antibody response against SARS-CoV-2 variants after several administrations of Bnt162b2
Clairon Q., **Prague M.**, Planas D., Bruel T., Hocqueloux L., Prazuck T., Schwartz O., Thiébaut R. and Guedj J. **PLOS Computational Biology** (IF=4.8) - 19(8) - e1011282- August 2023 (20 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Writing; Proofreading; Mentoring.

RI-23-4 Neutrophil Activation and Immune Thrombosis Profiles Persist in Convalescent COVID-19
Hocini H., Wiedemann A., Blengio F., ..., **Prague M.**, ..., and Lévy Y. **Journal of Clinical Immunology** (IF=8.2) - 43(5) - 882-893 - March 2023 (11 pages).
Contribution: Additional Statistical Analysis; Proofreading.

RI-23-3 Impact of non-pharmaceutical interventions, weather, vaccination, and variants on COVID-19 transmission across departments in France
Paireau J., Charpignon ML., Larrieu S., Calba C., Hozé N., Boëlle PY., Thiébaut R. **Prague M.**, and Cauchemez S. **BMC Infectious Diseases** (IF= 3.1) - 23 - 190 - March 2023 (12 pages).
Contribution: Additional Statistical Analysis; Proofreading; Mentoring.

RI-23-2 Leveraging Contact Network Information in Clustered Randomized Studies of Contagion Processes
Wang M.*, Staples P.*, **Prague M.**, Goyal R., DeGruttola V. and Onnela JP. **Observational studies** (IF= 1.2) - 9(2) - 157-175 - March 2023 (18 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Writing; Proofreading; Mentoring.

RI-23-1 Using a population-based Kalman estimator to model the COVID-19 epidemic in France: estimating associations between disease transmission and non-pharmaceutical interventions
Collin A., Hejblum B., Vignals C., Lehot L., Thiébaut R., Moireau P. and **Prague M.** **International Journal of Biostatistics** (IF= 1.83) - Epub ahead of print - Jan 2023 (18 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Writing; Proofreading.

RI-22-5 Estimation for dynamical systems using a population-based Kalman filter – Applications in computational biology
Collin A., **Prague M.** and Moireau P. **MathematicS in Action** (IF= 1.67) - 11(1) - 213-242 - Sept 2022 (29 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Writing; Proofreading

RI-22-4 Modelling the response to vaccine in non- human primates to define SARS-CoV-2 mechanistic correlates of protection
Alexandre M., Marlin R.*, **Prague M.***, Coléon S.,... and Thiébaut R. **eLife** (IF= 8.71) - 11 - e7542 - July 2022 (33 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Writing; Proofreading; Mentoring.

RI-22-3 Design, immunogenicity, and efficacy of a pan-sarbecovirus dendritic-cell targeting vaccine
Coléon S. Wiedemann A., Surénaud M., ..., **Prague M.**, ..., and Lévy Y. **eBioMedicine** (IF=

11.1) - 80(1) - 104062 - June 2022 (20 pages).
Contribution: Additional Statistical Analysis; Proofreading.

RI-22-2 Within-host models of SARS-CoV-2: What can it teach us on the biological factors driving virus pathogenesis and transmission?
Prague M., Alexandre M., Thiébaut R. and Guedj J. **Anaesthesia Critical Care & Pain medicine** (IF= 5.5) - 41(2) - 101055 - April 2022 (3 pages).
Contribution: Conceptualization of the Study; Writing; Proofreading; Mentoring.

RI-22-1 SAMBA: A novel method for fast automatic model building in nonlinear mixed-effects models
Prague M. and M. Lavielle **CPT: Pharmacometrics & Systems Pharmacology** (IF= 4.93 - 11(2) - 161-172 - Feb 2022 (11 pages).
Contribution: Statistical Analysis; Software Development; Writing; Proofreading.

RI-21-5 Robust and Efficient Optimization Using a Marquardt-Levenberg Algorithm with R Package marqLevAlg
Philipps V, Hejblum BP., **Prague M.**, Commenges D. and Proust-Lima C. **R Journal** (IF= 1.67) - 13(2) - 365-379 - Dec 2021 (14 pages).
Contribution: Software Development; Proofreading.

RI-21-4 Barrier Gesture Relaxation during Vaccination Campaign in France: Modelling Waning Immunity
Vignals, C., Dick, D. W., Thiébaut, R., Wittkop, L., **Prague M.**, and Heffernan, J. M. **MDPI Covid (viruses special issue)** (IF= 4.6) We got clearance from Inria to submit there but will not resubmit ever in MDPI journals as it is as date of 2023) - 1(2) -472-488 - Oct 2021 (16 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Writing; Proofreading; Mentoring.

RI-21-3 Targeting SARS-CoV-2 receptor-binding domain to cells expressing CD40 improves protection to infection in convalescent macaques
Marlin, R., Godot, V., Cardinaud, S., ..., **Prague M.**, ..., Le Grand R. **Nature Communication** (IF=16.6) - 12(1) - 5215 - Sept 2021 (9 pages).
Contribution: Additional Statistical Analysis; Proofreading.

RI-21-2 Between-group comparison of area under the curve in clinical trials with censored follow-up: Application to HIV therapeutic vaccines.
Alexandre M., **Prague M.**, Thiébaut R. **Statistical Methods in Medical Research** (IF= 2.3) - 30(9) - 2130-2147 - July 2021 (17 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Writing; Proofreading; Mentoring.

RI-21-1 Epidemioptim: A toolbox for the optimization of control policies in epidemiological models.
Colas, C., Hejblum, B., Rouillon, S., Thiébaut, R., Oudeyer, P. Y., Moulin-Frier, C., and **Prague M. Journal of Artificial Intelligence Research** (IF= 3.64) - 71(2) - 479-519 - July 2021 (40 pages).
Contribution: Conceptualization of the Study; Writing; Proofreading; Mentoring.

RI-20-3 Comparison of empirical and dynamic models for HIV viral load rebound after treatment interruption.
Bing A., Hu Y., **Prague M.**, Hill A., Li J., Bosch R., De gruttola V., and Wang R. **Statistical Communications in Infectious Diseases** (IF= 1.1) - 12(1) - 259-274 - Oct 2020 (6 pages).
Contribution: Development of Methodology; Proofreading; Mentoring.

RI-20-2 Machine learning and mechanistic modeling for prediction of metastatic relapse in early-stage breast cancer.
Nicolò, C., Périer, C., **Prague M.**, Bellera, C., MacGrogan, G., Saut, O., and Benzekry, S. **JCO Clinical Cancer Informatics** (IF= 1.2) - 4 - 259-274 - Sept 2020 (15 pages).
Contribution: Development of Methodology; Proofreading

RI-20-1 A model for establishment, maintenance and reactivation of the immune response after vaccination against Ebola virus.
Balelli, I., Pasin, C., **Prague M.**, Crauste, F., Van Effelterre, T., Bockstal, V., ... and Thiébaut, R. **Journal of Theoretical Biology** (IF= 2.32) - 495 - 110254 - March 2020 (20 pages).
Contribution: Development of Methodology; Writing; Proofreading; Mentoring.

RI-19-2 Properties and pitfalls of weighting as an alternative to multilevel multiple imputation in cluster randomized trials with missing binary outcomes under covariate-dependent missingness.
E. Turner, L. Yao, F. Li and **Prague M.**
Statistical Methods in Medical Research (IF= 2.38) - 29(5) - 1338-1353 - July 2019 (15 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Writing; Proofreading; Mentoring.

RI-19-1 Dynamics of the Humoral Immune Response to a Prime-Boost Ebola Vaccine: Quantification and Sources of Variation
C. Pasin, I. Balelli, T. Van Effelterre, V. Bockstal, L. Solforosi, **Prague M.**, M. Douoguih and R. Thiébaut
Journal of Virology (IF= 4.37) - 93(18) - e00579-19 - August. 2019 (10 pages).
Contribution: Proofreading; Mentoring.

RI-18-2 Adaptive protocols based on predictions from a mechanistic model of the effect of IL7 on CD4 counts
L. Vilain, D. Commenges, C. Pasin, **Prague M.** and R. Thiébaut
Statistics in Medicine (IF= 1.84) - 38(2) - 221-235 - Sept. 2018 (14 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Writing; Proofreading; Mentoring.

RI-18-1 High HIV-1 RNA Among Newly Diagnosed People in Botswana
V. Novitsky, **Prague M.**, Moyo S., Gaolathe T., MmalaneM., Kadima Yankinda E., Chakalisa U., Lebelonyane R., Khan N., Powis K., and others
AIDS research and human retroviruses (IF= 1.5) - 34(3) - 300-306 - Mars. 2018 (6 pages).
Contribution: Additional Statistical Analysis; Writing; Proofreading.

RI-17-5 CRTgeeDR: an R Package for Doubly Robust Generalized Estimating Equations Estimations in Cluster Randomized Trials with Missing Data.
Prague M., Wang R. and De Gruttola V
R journal (IF= 2.68) - 9(2) - 105-115 - Dec. 2017 (10 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Software Development; Writing; Proofreading; Mentoring.

RI-17-4 Universal test and treat and the HIV epidemic in rural South Africa: phase 4, open-label, community cluster randomised trial.
C. Iwuju, J. Orne-Gliemann, J. Larmarange, ..., **the TASP study group**
Lancet HIV (IF= 14.8) - 5(3) - 116-125 - Nov. 2017 (9 pages).
Contribution: Additional Statistical Analysis; Proofreading.

RI-17-3 Modeling CD4+ T cells dynamics in HIV-infected patients receiving repeated cycles of exogenous Interleukin 7
A. Jarne, D. Commenges, L. Villain, **M. Prague**, Y. Lévy, and R. Thiébaut
Annals of applied statistics (IF= 1.79) - 11(3) - 1593-1616 - Oct. 2017 (23 pages).
Contribution: Mentoring of PhD student, methodology.

RI-17-2 Review of Recent Methodological Developments in Group-Randomized Trials: Part 2—Analysis
L. Turner, F. Li, J. Gallis, **M. Prague**, and D. Murray
American Journal Of public Health (IF= 5.38) - 107(7) - 1078-1086 - June. 2017 (8 pages).
Contribution: Proofreading; Mentoring.

RI-17-1 Review of Recent Methodological Developments in Group-Randomized Trials: Part 1—Design
L. Turner, F. Li, J. Gallis, **M. Prague**, and D. Murray
American Journal Of public Health (IF= 5.38) - 107(6) - 907-915 - May. 2017 (8 pages).
Contribution: Proofreading; Mentoring.

RI-16-5 Point-of-care Cepheid Xpert HIV-1 Viral Load Test in Rural African Communities is Feasible and Reliable
S. Moyo, T. Mohammed, K. Wirth, **M. Prague**, K. Bennett, M. Holme, L. Mupfumi, P. Sebogodi, N. Moraka, C. Boleo, ..., E. Tchetgen Tchetgen, M. Essex, S. Lockman and V. Novitsky
Journal of clinical Microbiology (IF= 2.92) - 54(12) - 3050-3055 - Nov. 2016 (5 pages).
Contribution: Additional Statistical Analysis; Proofreading.

RI-16-4 Dynamic versus marginal structural models for estimating the effect of HAART on CD4 in observational studies: application to the Aquitaine Cohort study and the Swiss HIV Cohort Study.
M. Prague, Commenges D., Gran JM., Ledergerber B., Young J., Furrer H. and Thiébaut R.
Biometrics (IF=1.91) - 73(1) - 294-304 - July 2016 (10 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Software Development; Writing; Proofreading; Mentoring.

RI-16-3 Use of dynamical models for treatment optimization in HIV infected patients: a sequential Bayesian analysis approach
M. Prague
Journal de la statistique française - 157(2) - September 2016 (∼ 38 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Writing; Proofreading; Mentoring..

RI-16-2 Accounting for interference variables using semi-parametric augmentation for improving efficiency in clustered randomized trials with missing at random outcomes
M. Prague, Wang R., Stephens A., Tchetgen Tchetgen E and DeGruttola V.
Biometrics (IF=1.91) - 72(4) - 1066-1077 - April 2016 (11 pages)
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Software Development; Writing; Proofreading; Mentoring.

RI-16-1 Superior Efficacy of a Human Immunodeficiency Virus Vaccine Combined with Antiretroviral Prevention in Simian-Human Immunodeficiency Virus-Challenged Nonhuman Primates
Le Grand R., Bosquet N., Dispinseri S., Hopewell N., Gosse L., Desjardins D., Shen X., Tomaras G., Saidi H., **M. Prague**, Barnett S., Thiebaut R., Cope A., Scarlatti G., Shattock R.J.
Journal of Virology (IF=3.31) - 90(11) - 5315-5328 - March 2016 (14 pages).
Contribution: Additional Statistical Analysis; Proofreading.

RI-14-1 Quantifying and Predicting the Effect of exogenous Interleukin-7 on CD4+T cells in HIV-1 Infection.
Thiébaut R., Drylewicz J., **Prague M.**, Lacabaratz C., Beq S., Crough T., Sekaly R.P., Lederman M.M., Sereti I., Commenges D. and Levy Y.

PLOS comp. Biol. (IF=4.87) - 10(5) - e1003630 - May 2014 (12 pages).
Contribution: Statistical Analysis; Writing; Proofreading.

RI-13-2 Dynamical models of biomarkers and clinical progression for personalized medicine: the HIV context.
Prague M., Commenges D. and Thiébaut R.
Advanced Drug Delivery Review (IF=11.5) - 65(7) - 954-965 - June 2013 (12 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Software Development; Writing; Proofreading; Mentoring.

RI-13-1 NIMROD: A Program for Inference via Normal Approximation of the Posterior in Models with Random effects based on Ordinary Differential Equations.
Prague M., Commenges D., Guedj J., Drylewicz J. and Thiébaut R.
Computer methods and Programs in Biomedecine (IF=1.53) - 111(2) - 447-458 - June 2013 (12 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Software Development; Writing; Proofreading; Mentoring.

RI-12-2 Treatment monitoring of HIV infected patients based on mechanistic models.
Prague M., Commenges D., Drylewicz J. and Thiébaut R.
Biometrics (IF=1.83) - 68(3) - 902-911 - September 2012 (10 pages).
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Software Development; Writing; Proofreading; Mentoring.

RI-12-1 Acute versus chronic partial sleep deprivation in middle-aged people: differential effect on performance and sleepiness.
Philip P, Sagaspe P, **Prague M.**, Tassi P, Capelli A, Bioulac B, Commenges D, Taillard J.
Sleep (IF=5.05) - 35(7) - 997-1002 - September 2012 (6 pages).
Contribution: Additional Statistical Analysis; Proofreading.

1.3.4 Peer-reviewed national journals (2)

RN-23-1 Modéliser la COVID-19 : de la population à l'individu.
Vignals C., Hejblum B. and **Prague M.** **Interstice** Online - June 2023.
Contribution: Conceptualization of the Study; Writing; Proofreading; Mentoring.

RN-14-1 Modéles mathématiques dynamiques pour la médecine personnalisée.
Thiébaut R., **Prague M.**, Commenges D. **Medecine/Science ITMO Santé publique** 30(2) - 6-9 - Nov 2014 (3 pages).
Contribution: Proofreading.

1.3.5 Submitted papers with available preprints (6)

Sub-5 On the design of trials for the evaluation of HIV viral setpoint during analytical antiretroviral treatment interruptions
Alexandre* M., **M. Prague***, Lelièvre JD., Lhomme E., Richert L., Wittkop L., Lévy Y. and Thiébaut R.
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Writing; Proofreading; Mentoring.

Sub-4 Comparative assessment of methodologies to estimate NPI effectiveness in the COVID-19 context: A simulation study

Ganser I., Paireau J., D. Buckeridge, S. Cauchemez, Thiébaut R. and **M. Prague**
Contribution: Conceptualization of the Study; Development of Methodology; Writing; Proofreading; Mentoring.

Sub-3 Antibody response determinants to rVSV-ZEBOV-GP and Ad26.ZEBOV/MVA-BN-Filo Ebola vaccines: a modelling study from the PREVAC randomized trial
Valayer S., Alexandre M., **M. Prague**, Beayogui AH., ..., Richert L. and Lhomme E.
Contribution: Development of Methodology; Proofreading; Mentoring.

Sub-2 Viral rebound kinetics following single and combination immunotherapy for HIV/SIV
M. Prague, J. Gerold, I. Balelli, C. Pasin, J. Li,D. Barouch, J. Whitney, and A. Hill
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Software Development; Writing; Proofreading; Mentoring.

Sub-1 Effects of interventions and optimal strategies in the stochastic system approach to causality
D. Commenges and **M. Prague**
Contribution: Conceptualization of the Study; Development of Methodology; Statistical Analysis; Software Development; Writing; Proofreading; Mentoring.

1.3.6 Invited International conferences (11)

HCI*-23-3 Pharmacometric modelling to inform vaccine development.
Prague M.
World congress on basics and clinical pharmacology. , Glasgow UK, 2-7th July 2023. (Invited Speaker)

HCI*-23-1 Joint modeling of viral and humoral response in Non-human primates to define mechanistic correlates of protection for SARS-CoV-2.
Prague M., M. Alexandre, R. Marlin, Roger le Grand, Y. Levy and R. Thiébaut.
Society of mathematical Biology conference, Columbus USA, 16-21th July 2023. (Invited speaker)

HCI*-22-1 Elicitation of SARS-CoV-2 mechanistic correlates of protection using mechanistic models.
Prague M., Alexandre M., Marlin R, Le grand R, Levy Y and Thiébaut R.
The Canadian Applied and Industrial Mathematics Society, Online, 13-16th June 2022. (Invited speaker)

HCI*-21-5 Multi-level modeling of COVID-19 epidemic dynamics in French regions, estimating the combined effects of multiple non-pharmaceutical interventions.
Prague M., Hejblum B., Moireau P., Thiébaut R. and Collin A.
Society of mathematical Biology conference, Online, 13-17th June 2021. (Invited speaker)

HCI*-21-2 Viral dynamics as an outcome in HIV therapeutic vaccine trials: from AUC to dynamical modelling
Alexandre M., **Prague M.** and Thiébaut R.
CMStat , online, 18-20th Dec 2021 (Invited oral presentation by mentored PhD Student).

HCI*-21-1 Leveraging random effects to estimate the impact of non-pharmaceutical interventions on epidemic dynamics across French regions
Prague M., Collin A., Wittkop L., Dutartre D., Clairon Q., Moireau P., Thiébaut R. and Hejblum B.
Channel Network Conference, online, 7-9th April 2021 (Invited speaker).

HCI*-19-2 Evaluation of primary endpoint assessing HIV therapeutic vaccine efficacy during analytical treatment interruption studies.
Alexandre M., Thiébaut R., Levy Y. and **Prague M.**
4th workshop on virus dynamics, Paris, 21-23 Oct. 2019 (Invited oral presentation by mentored PhD student).

HCI*-19-1 Parameter estimation in nonlinear mixed effect models based on ordinary differential equations - an optimal control approach.
Clairon Q., Pasin C., Balelli I., Thiébaut R., and **Prague M.**
12th International Conference on Computational and Methodological Statistics, London, 14-16 Dec. 2019 (Invited oral presentation by mentored Postdoctoral fellow).

HCI*-17-1 Integrated approaches for analysis of cluster randomized trials. New development in analysis
Prague M.
Society for clinical trials, Liverpool, UK, 11-14 May 2017 (Invited speaker).

HCI*-16-1 Inverse-probability-weighted semi-parametric estimation of treatment effect in cluster randomized trials with missing data.
Prague M. and De Gruttola V.
Society for clinical trials, Montreal, Canada - May 15th-18th 2016 (Invited speaker).

HCI*-14-1 Comparison of GEE-based methods in cluster-randomized trial with missing data when outcome depend on other patients covariates.
Prague M., R. Wang, E. Tchetgen Tchetgen and V. De Gruttola
Joint Statistical Meeting, Boston, USA - August 2nd-7th 2014 (Invited speaker).

1.3.7 Competitive International conferences (12)

HCI-23-5 Effects of public health interventions against COVID-19 in France.
Ganser I., **Prague M.**, D. Buckridge and R. Thiébaut
Conference on Retroviruses and Opportunistic Infections CROI, Seattle, March 19-22th Feb 2023 (Poster in competitive conference).

HCI-23-4 Which endpoint to choose during antiretroviral treatment interruption?
Alexandre M., **Prague M.**, the VRI group and Thiébaut R.
Conference on Retroviruses and Opportunistic Infections CROI, Seattle, March 19-22th Feb 2023 (Poster in competitive conference).

HCI-23-2 Modeling neutralization capacities of Covid-19 vaccines.
Clairon Q., **Prague M.**, Thiebaut R. and Guedj J.
Population Approach Group in Europe Conference, A Curuna Spain, 28-30th June 2023. (Oral presentation by mentored Postdoctoral fellow a competitive conference)

HCI-22-3 SAMBA: a new algorithm for automatic construction of nonlinear mixed-effects models.
Prague M. and Lavielle M.
Population Approach Group in Europe Conference Ljubjana Slovenia, 28th June -7 July 2022. (Competitive conference).

HCI-22-2 SARS-CoV-2 mechanistic correlates of protection in non-human primates: insight from modelling response to vaccines.
Alexandre M., Marlin R., **Prague M.**, Thiébaut R. and Lévy Y.
Population Approach Group in Europe Conference, Ljubjana Slovenia, 28th June -7 July 2022. (Oral presentation by mentored PhD Student in a competitive conference)

HCI-21-4 Dynamics of the humoral immune response to a two-dose heterologous vaccine regimen against Ebola virus.
Alexandre M., **Prague M.** and Thiébaut R.
Population Approach Group in Europe Conference, Online, 2-7 Sept 2021. (Oral presentation by mentored PhD Student in a competitive conference)

HCI-21-3 Using population approach to model COVID-19 epidemics in France: estimating the burden of SARS-Cov-2 and the effects of non-pharmaceutical interventions.
Prague M., Hejblum B., Moireau P., Thiébaut R. and Collin A.,
Population Approach Group in Europe Conference, Online, 2-7 Sept 2021. (Speaker in a competitive conference)

HCI-20-1 Viral rebound kinetics following single and combination immunotherapy for HIV/SIV
Prague M., J. Gerold, I. Balelli, C. Pasin, J. Z. Li, D. Barouch, J. Whitney, A. L. Hill
Conference on Retroviruses and Opportunistic Infections CROI, Boston, 8-11th March 2020 (Poster in competitive conference).

HCI-18-3 Use of mathematical modeling for optimizing and adapting immunotherapy protocols in HIV-infected patients
Pasin C., Villain L., Dufour L., Commenges D., **Prague M.** and Thiébaut R.
Population Approach Group in Europe Conference Montreux, Switzerland, 28-30 May 2018 (Oral presentation by collaborator in competitive conference).

HCI-18-2 In silico clinical trials for evaluation of HIV short-cycle strategies.
Prague M., Pasin C., Thiébaut R. and the ANRS CO3 Study group
Conference on Retroviruses and Opportunistic Infections CROI, Boston, March 4th-7th 2018 (Poster in competitive conference).

HCI-18-1 HIV rebound kinetics following TLR7-agonist and therapeutic vaccine administration
Gerold J., Balelli I., Pasin C., Lim S., Barouch D., Whitney J., **Prague M.** and Hill A. L.
Conference on Retroviruses and Opportunistic Infections CROI, Boston, March 4th-7th 2018 (Poster in competitive conference).

HCI-13-1 From in vivo to in vitro quantification of antiretroviral drugs effects based on dynamical models of HIV.
Prague M., Commenges D. and Thiébaut R.
HIV Dynamics and evolution, Utrecht, Netherlands - May 8th-11th 2013 (Speaker in competitive conference).

1.3.8 Other International conferences (25)

CI-23-2 Using transcriptomic information in mechanistic models of immune response.
Thiébaut R., Hejblum B., Ba K. and **Prague M.**
6th Virus dynamics workshop, Nagoya Japan, 4-6th July 2023. (Oral presentation by collaborator).

CI-23-1 Use of priors in automated model building strategies for nonlinear mixed effects models.
Prague M. and Lavielle M.
International Society of biometry Conference, Milan Italy, 27-31th August 2023. (Speaker).

CI-21-3 Barrier gesture relaxation during vaccination campaign in France: modeling impact of waning immunity
Vignals C., Dick D., Thiébaut R, Wittkop L., **Prague M.** and Hefferman J.
Epidemics8, online, 1-3rd Dec. 2021. (Poster).

CI-21-2 A novel method for fast automatic model building in nonlinear mixed-effects models
Prague M. and Lavielle M.
5th Virus dynamics workshop, online, 4-6th Oct 2021. (Speaker).

CI-21-1 Accounting for time-dependant confounding variables in mechanistic ODE model: simulations and application to a vaccine trial
Alexandre M., **Prague M.** and Thiébaut R.
International Society of biometry Conference, online, 19-22th July 2021. (Oral presentation by mentored PhD student).

CI-20-1 Comparison of AUC in clinical trials with follow-up censoring: Application to HIV therapeutic vaccines
Alexandre M., Thiébaut R., Levy Y., **Prague M.**
International Society of biometry Conference, online, 23-27th August 2020. (Oral presentation by mentored PhD student).

CI-19-5 Viral rebound kinetics following single and combination immunotherapy for HIV/SIV
Prague M., Gerold J., Balelli I., Pasin C., Whithney J., Barouch D., Hill A. L.
4th workshop on virus dynamics, Paris, 21-23 Oct. 2019 Oral presentation by collaborator).

CI-19-4 A regularisation method for the problem of parameter estimation in ODE-mixed effect models: application to analysis of Ebola vaccine humoral response.
Clairon Q., Thiébaut R., and **Prague M.**
4th workshop on virus dynamics, Paris, 21-23 Oct. 2019 (Oral presentation by mentored Postdoctoral fellow).

CI-19-3 Machine learning versus mechanistic modeling for prediction of metastatic relapse in breast cancer.
Nicolò C., Périer C., **Prague M.**, MacGrogan G., Saut O., Benzekry S.
Population Approach Group in Europe Conference, Stockholm, 11-14 June 2019. (Oral presentation by collaborator).

CI-19-2 Modeling viral rebound in HIV therapeutic vaccine studies.
Alexandre M., Thiébaut R., and **Prague M.**
Population Approach Group in Europe conference, Stockholm, 11-14 Jun. 2019 (Poster by mentored PhD student).

CI-19-1 Evaluation of weighting and imputation methods to deal with missing outcomes in cluster randomized trials.
Turner L., Li F., and **Prague M.**
Society for clinical trials conference, New Orleans, 19-22 May 2019. (Oral presentation by collaborator).

CI-18-3 Performance of weighting as an alternative to multilevel multiple imputation in cluster randomized trials with missing binary outcomes, Developments in cluster randomised and stepped wedge designs.
Turner L., Li F. and **Prague M.**
Developments in cluster randomised and stepped wedge designs, London, 21-22 Nov. 2018. (Oral presentation by collaborator).

CI-18-2 Optimizing the administration of IL7
Villain L., Pasin C., **Prague M.** and Thiébaut R.
International Biometrics Society, Barcelona, Spain, 09-13 July 2018 (Oral presentation by mentored PhD student)

CI-18-1 Fitting pharmacokinetics data with a population-based Kalman filters.
Prague M. Collin A. and Moireau P.
International Biometrics Society, Barcelona, Spain, 09-13 July 2018 (Speaker).

CI-17-1 Modeling the humoral immune response to Ebola vaccine.
Pasin C., **Prague M.**, Eggo R., Van Effelterre T., Balelli I., ... and Thiébaut R
Systems Immunology and Vaccine Design, Heidelberg, Germany, 9-10 Oct. 2017. (Oral presentation by collaborator).

CI-16-2 Population Modelling by Examples II
Population Modeling Working Group including **M. Prague**
SummerSim conference, Canada - July 24-27th 2016 (including proceedings 8 pages)

CI-16-1 Estimating the Marginal Effect of Interventions to Reduce Spread of Communicable Diseases: What can be gained from Contact Network Information?
Prague M., Staples P., Onnela JP., Tchetgen Tchetgen E. and De Gruttola V.
ENAR, Austin Texas, USA - March 6th-9th 2016 (Speaker).

CI-15-2 Leveraging classical analysis of cluster randomized trials with contact network information infectious diseases.
Prague M., Staples P., Onnela JP. and De Gruttola V.
NIH Workshop Quantitative Methods and Models in the Era of Big Data, Washington DC, USA - Nov 9th-10th 2015(Poster).

CI-15-1 Accounting for informative missingness, interaction and interference in cluster randomized trials.
Prague M., Wang R., Stephens A., Tchetgen Tchetgen E. and DeGruttola V.
Society for Clinical Trials, Washington DC, Arlington, USA - May 17th-20th 2015 (Speaker).

CI-14-3 From descriptive to mechanistic models to study causal effects: application to the effect of HAART on CD4 count.
Prague M., Commenges D., Gran J.M., Aalen O. and Thiébaut R.
Joint Statistical Meeting, Boston, USA - August 2nd-7th 2014 (Speaker).

CI-14-2 Using mechanistic models to analyze the effect of interleukins 7 treatment in HIV infected patients
Jarne A., Thiébaut R., **Prague M.** and Commenges D.
International Biometric Society, Florence, Italy - July 6th-11th 2014. (Oral presentation by collaborator).

CI-14-1 Mechanistic versus marginal structural models for estimating the effect of HAART on CD4 counts
Commenges D., **Prague M.** and Thiébaut R.
Medical Research Council Conference on Biostatistics, Cambridge, UK - April 24th-26th 2014. (Oral presentation by collaborator).

CI-12-2 Toward information synthesis with mechanistic models of HIV dynamics.
Prague M., Commenges D. and Thiébaut R.
International society for Clinical Biostatistics, Bergen, Norway - August 21st-25th 2012 (Speaker).

CI-12-1 Bayesian MAP Estimation in Models with Random effects based on Ordinary Differential Equations applied to Treatment Monitoring in HIV.

Prague M. and Commenges D.
Eurandom Workshop on Parameter Estimation for Dynamical Systems (PEDS II), Eindhoven, Neetherland - June 4th-6th 2012 (Speaker).

CI-11-1 Treatment monitoring of HIV infected patients: optimal drug dose control.
Prague M., Commenges D., Drylewicz J. and Thiébaut R.
International Biometric Society, Bordeaux, France - April 11th-13th 2011 (Speaker).

1.3.9 Invited National Conferences (3)

CN*-23-1 Utilisation des données de transcriptomique pour informé les modèles mécanistes de réponse immunitaire.
Thiébaut R., Hejblum B., Ba K. and **Prague M.**
AC modélisation ANRS MIE, Paris, France. 24-25th Oct. 2023. (Invited oral presentation by collaborator)

CN*-21-3 Modeling B cells response.
Prague M.
AC modélisation ANRS MIE, Paris, France. 15 Oct 2021. (Invited Speaker).

CN*-15-1 Thesis Award MJLD: Use of dynamical models for treatment optimization in HIV infected patients.
Prague M.
JdS French Statistics Society conference, Lille, France, 1-5th June 2015. (Invited Speaker).

1.3.10 Other National conferences (14)

CN-23-1 High-dimension Mechanistic Model Building using LASSO Approaches : Application to Ebola Vaccination.
Gabaut A. **Prague M.**
GDR Statistiques et Santé, Toulouse, France. 16-17th Nov. 2023 (Oral presentation by mentored PhD Student).

CN-22-2 Modeling the temporal evolution of the neutralizing activity against SARS-CoV-2 variants after several administration of Bnt162b2.
Clairon C., Thiébaut R., Guedj J. and **Prague M.**
AC modélisation ANRS MIE, Bordeaux, France. 21-22th Nov. 2022. (Oral presentation by mentored Postdoctoral fellow).

CN-22-1 Estimation of the effect of non-pharmaceutical interventions and vaccination against COVID-19 in France using dynamical models.
Ganser I., Buckridge D., Thiébaut R. and **Prague M.**
AC modélisation ANRS MIE, Bordeaux, France. 21-22th Nov. 2022. (Oral presentation by mentored PhD Student).

CN-21-2 Méthode de comparaison d'aires sous la courbe dans des essais cliniques avec arrêt prématuré du suivi: application aux vaccins thérapeutiques contre le VIH.
Alexandre M., Thiébaut R. and **Prague M.**
JdS French Statistics Society conference, online, France, 1-6th June 2021.(Oral presentation by mentored PhD Student).

CN-21-1 Estimation paramétrique de modèles mixtes définis par des equations différentielles : une approache basée sur le contrôle optimal.
Clairon Q., Pasin C., Balelli I., Thiébaut R. and **Prague M.**
GDR Statistiques et Santé, online, France, 21-22th Oct. 2021.(Oral presentation by mentored Postdoctoral fellow).

CN-19-1 Modeling epidemics using networks.
Prague M., Staples P., De Gruttola V. and Onnela JP.
JdS French Statistics Society conference, Nancy, France, 3-7th June 2019. (Speaker).

CN-17-3 Use of mechanistic models for in Silico trials: Evaluating new strategies design for HAART in HIV infected patients.
Prague M., Commenges D. and Thiébaut R.
GDR Statistiques et Santé, Bordeaux, France, 9-10th Oct. 2017. (Speaker).

CN-17-2 Non linear mixed effect models based on Ordinary Differential equations.
Prague M., Commenges D. and Thiébaut R.
GDR Mamovi, Lyon, France, 27-28th Sept. 2017. (Speaker).

CN-17-1 Joint-state and parameters estimation using Kalman-based filters.
Prague M., Collin A. and Moireau P.
JdS French Statistics Society conference, Avignon, France, 1-4th June 2017. (Speaker).

CN-16-1 Estimation doublement robuste de l'effet marginal d'intervention pour les essais randomisées en cluster.
Prague M., Tchetgen Tchetgen E. and De Gruttola V.
JdS French Statistics Society conference, Montpellier, France, 1-4th June 2016. (Speaker).

CN-13-1 Inférence par Approximation Normale de l'a posteriori dans les modèles dynamiques à Effets mixtes
Prague M., Commenges D., Guedj J., Drylewicz J. and Thiébaut R.
JdS French Statistics Society conference, Toulouse, France, 21-31th May 2013. (Speaker).

CN-12-2 Illustration of information synthesis of clinical trials with mechanistic models of HIV dynamics
Prague M., Commenges D. and Thiébaut R.
GDR Statistiques et Santé, Rennes, France, 20 Sept. 2012. (Speaker).

CN-12-1 R Package "marqLevAlg" : the Marcquardt-Levenberg algorithm an alternative to "optimx"
Prague M., Diakité A. and Commenges D.
UseR France, Bordeaux, France, 2-3 July 2012. (Speaker).

CN-11-1 Estimation in Differential Equations and prediction of treatment response in HIV infected patients
Prague M. and Commenges D.
GDR Statistiques et Santé, Paris, France, 30 May 2011. (Speaker).

1.3.11 Invited Seminars - selection in external institutions (15)

S-23-1 Defining Mechanistic correlates of protection
MRC Cambridge, Biostatistics department, online, 24 May 2023.

S-22-1 SARS-CoV-2 mechanistic correlates of protection: insight from modeling response to vaccine
York university, in-host seminar series, online, 14 oct. 2022.

S-21-1 Evaluating the longevity of humoral response: from Ebola to SARS-CoV-2
EpidemiOptim: a toolbox for the optimization of control policies in epidemiological models
Google DeepMind seminar, online, 7 Janv. 2021.

S-20-2 Evaluating the longevity of humoral response: from Ebola to SARS-CoV-2
Modcov19 Seminar series & Health data Hub, online, 1 April and 4 Nov. 2020.

S-20-1 Multi-level modeling of early COVID-19 epidemic dynamics in French regions and estimation of the lockdown impact on infection rate
European Network for Business and Industrial Statistics, online, 6 July 2020.

S-19-1 Updates on estimation of parameters in dynamical models
Harvard Program for evolutionary dynamics, Boston, USA, 5 Feb. 2019.

S-18-1 Use of mechanistic models for in silico trials
Institut Gustave Roussy, Biostat/Oncostat, Paris, France, 12 Feb. 2018.

S-17-2 Use of mechanistic models for in Silico trials
Harvard School of Public Health, Biostatistics Department, Boston USA, 10 Apr. 2017.

S-17-1 Doubly Robust estimators in Cluster Randomized Trials
University of Pennsylvania, Biostatistics Department, Philadelphie, USA, 3 Apr. 2017.

S-15-2 Estimating intervention effect in cluster randomized trials: handling, missing data, interferences
IRD, MERIT Team, Paris, France, 15 Dec. 2015.

S-15-1 Double robust estimation of causal effect of intervention in cluster randomized trial with missing data
Inserm U1219, Department of biostatistics, Bordeaux, France, 12 Mars 2015.

S-13-3 Estimation of HAART treatment effect in observational studies
University of Oslo, Department of statistics, University of Oslo, Norway, 10 Dec. 2013.

S-13-2 Utilisation de modèles dynamiques pour l'optimisation des traitements des patients infectés VIH
Inserm U1137, University of Paris Diderot, Paris, France, 27 Nov. 2013.

S-13-1 Dynamical models in the HIV context: Prediction of treatment responses and optimization
John Hopkins hospital, Pharmacological sciences Department., Baltimore, USA, 26 Oct. 2013.

S-12-1 Estimation and treatment optimization in HIV infected patients
University of Liège, Department of methodological statistics, Liège, Belgium, 11-13 Dec. 2012.

1.3.12 Softwares (6)

Soft-23-1 R package "Rsmlx" is a library that allows to interface optimization algorithms with models of epidemic propagation. Epidemiological models are wrapped in OpenAI Gym interfaces, making them readily compatible with state-of-the-art optimization algorithms.
Contribution: My contribution to this code is modest; nevertheless, I played an active role in the development of ideas surrounding the construction of random effects models and the integration of priors.

Soft-21-1 Python Toolboox "EpidemiOptim" is a library that allows to interface optimization algorithms with models of epidemic propagation. Epidemiological models are wrapped in OpenAI Gym interfaces, making them readily compatible with state-of-the-art optimization algorithms.
Contribution: I am one of the principal contributor to the concept underlying this code. I have been a tester for this code.

Soft-20-2 R package "SEIRcovid19FR" is an R package for multi-level inference of epidemic dynamics applied to COVID-19 in France.
Contribution: I am the principal contributor to the concept underlying this code. I contributed to the development of this code

Soft-20-1 R package "marqLevAlg" This algorithm provides a numerical solution to the problem of unconstrained local minimization (or maximization). It is particularly suited for complex problems and more efficient than the Gauss-Newton-like algorithm when starting from points very far from the final minimum (or maximum). Each iteration is parallelized and convergence relies on a stringent stopping criterion based on the first and second derivatives.
Contribution: I was the main developer of the first version that was later re-factored for publication by V. Philipps.

Soft-16-1 R package "CRTgeeDR" implements a semi-parametric GEE (Generalized Estimating Equation) estimator accounting for missing data with Inverse-probability weighting (IPW) and for imbalance in covariates with augmentation appraoches (AUG). The estimator IPW-AUG-GEE is Doubly robust (DR).
Contribution: I am the main developer and maintainer.

Soft-13-1 **The "NIMROD" software** is a Fortran program (Normal approximation Inference in Models with Random effects based on Ordinary Differential equations) devoted to the estimation in Ordinary Differential Equations (ODE) models with random effects. Although maximum likelihood based approaches are valuable options, both numerical and identifiability issues favor a Bayesian approach which can incorporate prior knowledge in a flexible way. NIMROD estimations relies on a normal approximation of the posterior that can be obtained by computing the maximum of the posterior distribution (MAP) by maximizing a penalized likelihood. NIMROD is now deprecated.
Contribution: I was the main developer.

1.3.13 Press Releases (18)

During the COVID-19 pandemic, a number of our works were highlighted by the media. Below is a non-exhaustive list of press releases where our work was cited:

- La Croix 05/05/2020 "Coronavirus l'intrigante géographie contagion.",
- Parisien 08/06/2020 "Covid-19 les manifestations risquent-elles d'accroitre la propagation du virus",
- Sud-Ouest 17/11/2020 "Covid-19 comment travaillent les épidemiologistes",
- France 2 journal TV 13h 29/12/2020,
- Le Parisien 06/01/2021 "Covid-19 l'ouest de la France restera-t-il epargné?",
- Le Parisien 29/04/2021 "Déconfinement à partir du 3 mai est-ce bien raisonnable d'un point de vue sanitaire",
- Le Parisien 06/09/2021 "Covid-19 y a-t-il vraiment une correlation entre le taux d'incidence et la vaccination",
- La tribune 19/07/2021 "Intelligence artificielle et épidemiologie deux clefs pour la santé publique",
- Le Monde 28/07/2021 "Covid-19 face à la quatrième vague les effets trop tardifs de l'accéleration de la vaccination",
- France 2 Journal Télévisé 20h 01/09/2021,

- Le monde 25/09/2021 "Covid-19 malgré cinq semaines de baisse continue en métropole les épidemiologistes restent prudents",
- La voix du Nord 4/11/2021 "Covid-19 jusqu'à quand faudra-t-il porter un masque?",
- Sud Ouest 04/01/2022 "Omicron avec une possible immunité collective le gouvernement amorce une bonne stratégie de santé publique",
- Sud Ouest 04/01/2022 "Levée des restrictions sanitaires il serait judicieux de porter le masque encore un peu",
- Sud-Ouest 06/02/2024 "Covid-19 une étude met en lumière les milliers de vies sauvées par les confinements et la vaccination",
- L'express 07/02/2024 "Covid-19 le confinement et les vaccins ont-ils été efficaces ce que révèle une étude",
- Le monde 08/02/2024 "Covid-19 sans la vaccination le nombre de morts aurait été le double en France",
- La Dépêche 09/02/2024 "Covid-19, les données prédisent 159000 décès supplémentaires une étude mesure l'impact du confinement et du vaccin en France.

Chapter 2

Statistical Methodology for Estimations and Building of Mechanistic Models

2.1 Bibliometry

The first research axis is predominantly theoretical, focusing on the complexities of solving inverse problems in nonlinear mixed-effect models (NLME), of efficiently predicting individual outcomes and, of optimizing input of the systems. This research mainly explores optimization techniques and strategies to construct and use robust and adequate models. The outputs of this axis are summarized in Table 2.1. **This theme gave birth to 9 articles in peer-reviewed international journals, 14 international conferences, 10 national conferences, 3 seminars and 3 softwares. It is also funded by 3 international grants and 2 national grants.** The interrelation of all these works is explained in the next subsections.

2.2 Mechanistic Model Definition

This research direction is the most theoretical aspect of my work. It is grounded in my biostatistics background and driven by the challenges encountered in processing data from both international and national projects, including collaborations with the Vaccine Research Institute. In most study in my work, we handle **longitudinal data collected for multiple individuals**. The poor performance of individual data fitting has been extensively studied since the 1970s. Lewis Sheiner highlighted the inefficiency of individual data fitting for parameter estimation, as it doesn't differentiate between between-subjects and within-subject variability, necessitating frequent data collection in each patient [180, 179]. It was also shown that biased conclusion on treatment effects can be derived from patient-by-patient fitting [178]. Thus, we need to develop efficient tool to model jointly repeated measurements taken for each patient.

Peer-reviewed International Journals		**First**	**Second**	**Before Last or Last**	**Other**
CPT: Pharmacometrics & Syst. Pharma.		RI-22-1			
Biometrics		RI-16-4			
R Journal					RI-21-5
MathematicS in Action			RI-22-5		
Computational meth. and prog. Biomed.		RI-13-1			
Computational Statistics				RI-23-6	
Journal Clinical Oncology					RI-20-2
Statistical Comm. Infectious Diseases					RI-20-3
French Statistical Society Journal		RI-16-3			
Total (9)		4	1	1	3
International conferences	**Invited**	**Speaker**	**Mentee**	**Poster**	**Other**
PAGE					HCI-22-3; CI-19-3
CMStat	HCI*-19-1				
ISBC		CI-23-1	CI-12-2		
Workshop on Virus Dynamics		CI-21-2	CI-19-4		CI-23-2
IBC		CI-18-1; CI-11-1			
Joint Statistical Meeting		CI-14-3			
Other		CI-12-1			CI-16-2; CI-14-1
Total (14)	1	6	2	0	5
National conferences	**Invited**	**Speaker**	**Mentee**	**Poster**	**Other**
GDR Stat & Santé ou Mamovi		CN-17-2; CN-13-1; CN-12-2; CN-11-1	CN-23-1; CN-21-1		
ANRS MIE AC					CN*-23-1
JdS SFdS	CN*-15-1	CN-17-1			
useR!		CN-12-1			
Total (10)	1	6	2	0	1
Seminars (3)	S-19-1; S-13-3; S-12-1				
International Grants (3)	IG-20-1; IG-19-1; IG-18-2				
National Grants (2)	NG-22-1; NG-17-1				
Sofwares (3)	Soft-23-1; Soft-20-1; Soft-13-1				

Table 2.1 – Bibliometry for axis "Statistical methodology for estimations and building of mechanistic models"

Nonlinear mixed-effects models (NLME) are particularly adapted for this purpose, which makes them the primary tool employed in my research. NLME models are a statistical approach used to analyze data that is non-linear in nature and includes both fixed and random effects. The **fixed effects** represent the population-average effects and are consistent across all individuals, possibly depending on measured explanatory covariates. **Random effects** account for individual-specific variations. These effects allow the model to account for variability in the data that is not explained by the fixed effects alone, which is the extra heterogeneity not captured by individual covariates. The model accounts for non-linear relationships between variables. In my research, this is typically achieved through **Ordinary differential equations** (ODE) which may or may not admit an analytic solution. Of note, few examples could be extended for stochastic differential equations or partial differential equations.

NLME are also referred to as mechanistic models. **Mechanistic models**, in a general scientific context, refer to models that are based on the underlying mechanisms. One of the key characteristic of mechanistic models is that they are a process-based representation of a phenomenon based on fundamental principles of kinetics. This is particularly useful in my research as we model biological processes related to immunology and virology to derive the ODE. Mechanistic models can become very complex, especially when they aim to describe intricate biological systems. This often requires significant computational resources for simulation and analysis.

We make the choice to present mechanistic model by dividing the description into three interconnected components: **the mathematical model, the statistical model, and the observation model**. Each part plays a crucial role in the modeling process, from representing biological processes to fitting the model to empirical data.

- **Mathematical Model (or Process Model):** The mathematical model is the core part that describes the underlying biological or physical processes, in our setting, using a system of ODE ($f(\psi, X, t)$). It consists of equations that represent the change in state of various components X (e.g., concentrations of different substances, populations of cells) over time t. The equations are typically derived based on biological knowledge and fundamental principles but could also be built by data-driven approaches (see Section 2.5). Parameters in the mathematical model ψ, like rate constants or initial conditions, are often biologically interpretable but not directly observable.

- **Statistical Model (or Parameter Model):** The statistical model deals with the uncertainty and variability in the parameters of the mathematical model. It includes prior distributions for the parameters, reflecting any previous knowledge or assumptions about their values. The statistical model also handles the variability between subjects (inter-individual variability rep-

resented by random effect η_i) or within the same subject over time (intra-individual variability represented by fixed effects ψ_0 and β for effect of covariates Z_i). It is crucial for parameter estimation, especially when fitting the model to data, and for quantifying the uncertainty in the estimates. For individual $i = 1, \dots N$, the individual parameters write, up to a transformation function h, $h(\psi_i) = h(\psi_0) + \beta Z_i + \eta_i$. The random effects η_i are normally distributed with mean 0 and variance Ω. The function h is often chosen, though not exclusively, to be log-normal because it constrains the diffusion rates of the ODE to be positive.

- **Observation Model (or Data Model):** The observation model connects the mathematical model to the real-world data. It describes how the actual data measurements are generated from the theoretical values predicted by the mathematical model. This often involves accounting for measurement errors or other forms of noise. The model may include factors like the sensitivity and specificity of the measurement techniques, sampling times, and the nature of the data (e.g., continuous, count data, binary outcomes). In particular, it is essential for translating between the scale and units of the mathematical model and the observed data. Finally, observation model may only represent a fraction of the compartments within the mechanistic model, resulting in incomplete information about all the components of the process. The observation Y_{ij} for individual i at time t_{ij}, with $j = 1, \dots, n_i$, is related with ODE outputs up to a transformation g such that $Y_{ij} = g(X(\psi_i, t_{ij})) + g_e(\xi, X(\psi_i, t_{ij}))\epsilon_{ij}$. Where $g_e(\xi, X(\psi_i, t_{ij}))$ is the error function of the marker and ϵ_{ij} is normally distributed of mean 0 and variance 1. Of note, in a constant error model, often referred to as white noise, $g_e(\xi, X(\psi_i, t_{ij}))$ is assumed to be a diagonal matrix.

In this mechanistic model $\mathcal{M}$, we are interested in estimation of the joint distribution of $\theta = (\psi_0, \beta, \Omega, \xi)$. Then for prediction, we are interested in the individual values of parameters ψ_i as well as their related uncertainty.

2.3 Inference in Mechanistic Model

Possessing the capability to manipulate, compare, and construct estimation approaches for addressing the inverse problem is crucial; this involves being able to perform forward simulation using mechanistic models. For forward simulation of ODE, we use standard numerical methods ranging from Euler's method and Runge-Kutta Methods to stiff ODE Solvers such as the Backward Differentiation Formula method (BDF) [38]. One of the interesting feature that make our problem difficult is the stiffness of many of our ODE. A **stiff ODE** system is one in which there is a significant disparity in the time scales of the processes or phenomena described by the ODE. In simpler terms, stiffness occurs

when some components of the system evolve much more rapidly than others, leading to numerical challenges when simulating the ODE. In my research to date, I have assumed that forward simulation methods exist and are reliable. Thus, I am mostly interested in inferring the parameters of a model based on available data, i.e. **solving the inverse problem**.

Another aspect of mechanistic model that is important to check before trying inference on the model is **identifiability**. This term refers to the ability to uniquely determine the model's parameters (or some of them) based on observed data. My research does not yet feature new development in this aspect. However, this is something we carefully check before attempting any inference on a given problem. First, **structural identifiability** relates to whether the model structure itself allows for unique estimation of parameters. It considers whether the mathematical relationships between model variables and parameters are such that no parameter can be changed without affecting the model's predictions. To assess this problem, we routinely use Differential Algebra for Identifiability in Systems, historically DAISY [17] and more recently a method that also ensures that derivatives of parameters are identifiable [51]. Second, **practical identifiability** considers whether, in practice, with a specific dataset and measurement noise, the parameters can be estimated with sufficient precision [114]. To assess this problem, we routinely use sensitivity analysis, Monte Carlo simulations and profile likelihood analysis. See Section 2.6 for details on future envisioned research on identifiability and optimal design.

The challenge is the development of population-based estimation approaches (CI-16-2). Maximum Likelihood Estimation is a commonly used method for estimating parameters in mechanistic models. It seeks to find the parameter values that maximize the likelihood of observing the observed data (y) given the model, which is expressed as the product of the probability density (or likelihood) of the observed data points [153]. The maximum likelihood estimator $\hat{\theta} = argmax\,[l(\theta, y)]$, where the likelihood can be written as:

$$l(\theta, y) = \prod_{i=1,\dots N} \int p(y_i|\eta_i, \theta) p(\eta_i, \theta) d\eta_i$$

However, because our methods often lack of identifiability, we initially use a Bayesian framework in which it is possible to input biological information on parameters values. In Bayesian modeling, you start with a prior distribution for the parameters and update it with the likelihood of the data to obtain the posterior distribution. Markov Chain Monte Carlo (MCMC), Variational Inference and importance sampling are common techniques used for Bayesian parameter estimation [125, 160, 162]. In this regard, my research initially focused on methods involving **penalized likelihood maximization** for NLME models (CI-11-1; CI-12-1; CN-17-2; CN-13-1; CN-11-1; S-12-1). The relationship between penalized likelihood estimation and Bayesian statistics lies in the incorporation of additional

information or constraints on the parameters. The algorithm maximizes $\log(l(\theta, y)) - J(\theta)$, where $J(\theta)$ increases as the parameter estimates deviate further from their prior values. Our algorithm uses the Levenberg-Marquardt gradient descent algorithm (RI-21-5; CN-12-1, released in an R package Soft-20-1), to approximate and maximize the penalized likelihood. The descent process is described as follows:

$$\theta_{k+1} = \theta_k - (H + \nu I)^{-1}(\theta_k)U(\theta_k),$$

where U and H represent the score and the Hessian of the likelihood, respectively, while ν is a term added to ensure the Hessian remains positive definite and approaches zero over iterations. The Hessian is approximated by a function of the scores:

$$G(\theta_k) = \sum_i U_i(\theta_k)U_i'(\theta_k) + \frac{\nu}{n}U(\theta_k)U'(\theta_k).$$

The score for each individual $i = 1, \ldots, N$ is expressed as:

$$U_i(\theta_k) = p(y_i|\theta)^{-1} \int p(y_i|\eta_i, \theta)\frac{d\log\left(p(y_i|\eta_i, \theta)\right)}{d\eta_i}d\eta_i,$$

which is evaluated using sensitivity equations of the ODE, adaptive Gaussian quadrature, and the Livermore Solver for Ordinary Differential Equations. This algorithm was implemented in a Fortran program Soft-13-1). Additionaly, an interesting feature of this research is the development of a rigorous stopping criterion defined as the relative distance to maximum (RDM):

$$\frac{1}{\dim(\theta_k)}U'(\theta_k)G^{-1}(\theta_k)U(\theta_k),$$

which serves as a proxy for the ratio between the numerical error and the statistical error, with a lower value indicating better performance (RI-13-1). The whole phD thesis on this topic as well as applications developed in it has been awarded by the SFdS PhD thesis award for applied statistics (CN*-15-1).

The calculation burden due to the complexity of ODE models, led us to progressively shift toward algorithms which does not require as many evaluations of the likelihood as a gradient-descent method. The **Stochastic Approximation of the Expectation-Maximization (SAEM) algorithm** is an iterative optimization method used for estimating the parameters of mechanistic models. It combines elements of both the Expectation-Maximization (EM) algorithm and stochastic approximation [111] extended for left-censored data [172]. In the E-step, SAEM computes the expected value (or conditional expectation) of the complete likelihood function given the current parameter estimates and observed data. To do so, it uses a Monte Carlo approach to approximate the expected likelihood,

often referred to as the population likelihood. Then, in the M-step, SAEM updates the parameter estimates to maximize the expected complete likelihood obtained in the E-step. SAEM incorporates a stochastic component by using small perturbations or noise in parameter updates. This helps escape local maxima and improve the convergence of the algorithm. SAEM iteratively alternates between the E-step and the M-step until convergence criteria are met. An evolution of this algorithm is the Mixture SAEM, see [115] for all details, it writes and iterates as follows for iteration k:

- S-step: Sample the latent variable $\psi_i^{(k)}$ according to the conditional distribution $p(.|y, \theta_k)$.
- E-step: Compute $H(y, \psi_i^{(k)}; \theta_k) = E(S(y, \psi_i^{(k)}, z)|y, \psi_i^{(k)}, \theta_k)$, where S is a function of the minimal sufficient statistics of the log-likelihood.
- AE-step: Update $Q_k(\theta)$ such that $Q_{k+1}(\theta) = Q_k(\theta) + \gamma_k(H(y, \psi_i^{(k)}, \theta_k) - Q_k(\theta))$, where γ_k is a decreasing sequence tuning the convergence rate of the algorithm. We usually take $\gamma_k = 1$ during an exploratory phase and $\gamma_k = 1/k$ in the smoothing phase.
- M-step: θ_{k+1} maximizes $Q_{k+1}(\theta)$.

Convergence is typically assessed based on changes and stabilization in parameter estimates and the complete likelihood. It is implemented in the software Monolix that we routinely use [116] (as disseminated in educational seminar S-19-1). We currently entertain relationship with the Simulation Plus Lixoft company owning this software. However, given Monolix's proprietary nature, which limits rapid source code evolution, and its inability to meet all the specific requirements for our applications, we continue the development of alternative estimation methods.

In this previous approach, the relationship that exists between parameters and time need to be parametrically defined. Semi-parametric models do not assume a specific functional form and are more flexible, allowing the data to dictate the structure of the model. Methods based on splines smoothing have been developed to fit these models [204]. However, they may also prove to be computationally demanding and the number of hyperparameters to fix (eg. splines basis, knots...) makes them very sensitive to parametrization [206]. We thus decided to develop an alternative semi-parametric approach for estimation in mechanistic model. **Data assimilation** methods are often used in environmental sciences and confronted with the complexity burden associated with large-scale systems [9]. Famous data assimilation approaches are based on **Kalman filters**. The Kalman filter is an algorithm that provides efficient computational recursive means to estimate the state of a process in a way that minimizes the mean of the squared error. The Unscented Kalman Filter particularly addresses some limitations of the traditional Kalman filter when dealing with nonlinear systems [102]. Together with collaborators (Annabelle Collin Inria Bordeaux Team Monc and Philippe Moireau Inria Team M3disim Paris Saclay), we extended Kalman filters to population approach (RI-22-5; CI-18-1;

CN-17-1). The potential curse of dimensionality caused by the population framework is then limited by covariance reduction techniques such as those proposed in the Reduced-Order Unscented Kalman Filter [138]. Note that this is not the first time that Kalman-based approaches have been used in mixed-effects strategies [45, 108, 196]. However, in this literature, the extended Kalman filters are used only to approximate the individual probability distribution function, whereas in our work we use the Kalman approach also for estimation at the population level. We investigated the use of this new fitting strategy on a large range of scenarios and we found that this method requires a large amount of data with a very frequent sampling to provide reliable estimates. Thus, we applied it to estimation of effects of non-pharmaceutical interventions in COVID-19 epidemics for which time series of cases and hospitalization are observed daily on a long course (see Section 4.3.1). See Section 2.6 for details on how we envision to use this method for sparse data.

The approaches based on the estimation of a given ODE model face the problem of not taking into account the presence of model misspecification. First, most of the time, the true initial conditions are unknown, which implies either assumptions on their values [117] or their estimation [89]. Second, they can face accuracy degradation when the inverse problem of parameter estimation is ill-posed due to practical identifiability issues [61]. Another method that we investigated is grounded in **optimal control theory**. It consists in optimization that deals with finding a control law for a dynamical system over a period of time such that an objective function is optimized. In a few words, the idea is to fit the ODE model to data together with a control variable that can be manipulated to influence the behavior of the system. It aims to regularize the estimation problem in the presence of model misspecification and practical identifiability issues, while avoiding the need to know or estimate initial conditions as nuisance parameters. Together with a postdoctoral fellow Quentin Clairon, we extended existing approach [29] to population approach. See Section 6.1 for the full version of the article presenting this work. Compared to the maximum likelihood method, we show through simulation examples that our method improves the estimation accuracy in possibly partially observed systems with unknown initial conditions or poorly identifiable parameters with or without model error (RI-23-6; HCI*-19-1; CI-19-4; CN-21-1). See Section 2.6 for details on how we believe this method would help us to assess model misspecification.

2.4 Validity of the Approach: Causality and Predictions

While it is feasible to manipulate mechanistic models, it is imperative to be convinced of their superiority for modeling available data. This necessitates a **comparative analysis with other available tools**. Even if the answer may be problem-specific, we postulate that mechanistic models may be superior in term of prediction abilities as they integrate biological knowledge. First, together with

collaborators (Rui Wang and Alison Hill at Harvard university), we focus on comparing two different modeling approaches for analyzing HIV-1 RNA viral load trajectories after antiretroviral treatment interruption (RI-21-2; CI-19-2). The models of interest are **splines** [210] and mechanistic models [149]. Both models aim to predict features of viral rebound, like viral set points and delay in rebound, and identify factors influencing these features. The study finds that both models offer reasonable fits to the data. The mechanistic models however lead to slightly more efficiency. This is even amplified in presence of limit of detection for viral load makers and lost of follow-up of patients due to antiretroviral resumption (RI-20-3). Second, together with collaborator (Sebastien Benzeckry Inria-Inserm Team Compo Marseille), we focus on evaluating predictive models for metastatic relapse in patients with early-stage breast cancer. It compares the effectiveness of **machine learning algorithms** [213] and mechanistic models [94] in predicting the risk of metastasis after surgical intervention. The model's predictive performances are comparable. However, we show that the mechanistic model provides estimates of the invisible metastatic burden at diagnosis and simulates metastatic growth. It can serve as a personalized prediction tool for managing patients with breast cancer. See Section 2.6 for details on how we believe statistical learning and deep learning methods can be adapted to address predictions and inference problems in our mechanistic models.

The goal of mechanistic models is often to **evaluate the effects of treatments or exposures over time**. This question may arise in a randomized but also in an observational setting. In the latter, causal problems may arise [25]. A **time-varying confounder** is a variable that influences, and is influenced by, the exposure (or treatment), while also being a risk factor for the outcome. This is for example the case of CD4 count when evaluating the effect of antiretroviral treatment. CD4 count is both the outcome and a covariate explaining treatment assignment. This dynamic nature poses significant challenges in accurately assessing causal relationships. In this setting, it has been demonstrated that a conventional regression analysis leads to biased estimates of the treatment effect, typically underestimating it, and may (wrongly) indicate a negative effect. This is called confounding by indication [203]. Marginal structural models (MSM) [168] have been proposed for dealing with this issue; this is based on choosing a causal model in terms of potential responses, which are often counterfactual, to the different treatment histories. The parameters of a MSM can be estimated through a weighted approach but other methods exist such as targeted maximum likelihood [152]. The weights are the inverse probability of treatment assignment and are obtained through a "treatment model" which includes the covariates linked to the outcome. Because data are correlated, we use an inverse probability weighted generalized estimating equation (GEE), see Section 4.2 for use of this tool in the setting of missing data. This approach has been applied for estimating the effect of antiretroviral on CD4 count [83, 31]. An alternative view to causality that does not use the potential responses representation is to use dynamic models. Among others, Bayesian decision

analysis [43], graphical models [49], dynamical models based on stochastic processes [33] and linear increment models [86] were pioneering as alternative approaches. Then assumptions needed for a causal interpretation of dynamic models have been presented [8, 34]. In a nutshell, causality operates over time and aligns with a mechanistic or system view since time is an intrinsic component through ODE [1]. We investigated and compared in simulation and real data all these approaches in a study interested in informative treatment assignment in observational studies (RI-16-4; CI-14-3; CI-14-1; S-13-3). We focus on developing dynamic models to estimate the effects of antiretroviral on CD4 counts in HIV patients, using data from Swiss HIV cohort. We conclude that mechanistic models provide a more accurate and more efficient estimation of antiretroviral's effect on CD4 counts in HIV patients than others. It is thus a valid tool for **causal inference**. We also demonstrate the superiority of mechanistic models in capturing the complexity of biological systems and accessing unmeasured information. See Section 6.2 for the full version of the article presenting this work. Because these models are based on the mechanisms of the system, they can be used to make **predictions** about the system's behavior under new conditions or in response to interventions, i.e. counterfactual to the actual study conditions. This is particularly useful in drug development and understanding individual disease progression. This will be further expanded in Section 3.2.

2.5 Model Building Strategies

A variety of models exist to fit data, and even with parallel computing, the adjustment of mechanistic models can be time-intensive. This is particularly true as the number of parameters or the size of the dataset grows. Typically, the evaluation of the most suitable model for a specific dataset involves **optimizing an information criterion** to determine the best fit. It could be Akaike Information Criterion (AIC), Bayesian Information Criteria (BIC), or corrected BIC which is most suited for NLME models (BICc) [46]. These criteria balance model fit with complexity but do not explicitly penalize model complexity sufficiently in all scenarios, potentially leading to overfitting. Other metrics based on external validation, bootstrapping/leave-one-out and predictive abilities could also be considered and expanded [20, 35].

Having a large number of individuals from multiple studies or incorporating information from different sources of information improves practical identifiability of mechanistic models [80]. However, it increases computation times and may make the estimation intractable; an alternative approach is to use a **Sequential Bayesian Analysis** [207] (SBA). SBA is a statistical method in which data is analyzed in stages, and the results from each stage inform the analysis of subsequent data. This approach updates beliefs or estimates in light of new data, using Bayes' theorem. It has been shown that for most generalized linear and non-linear equations, the posteriors are consistent and admits a

normal approximation [74]. In each step, the posterior distribution from the previous stage becomes the prior for the current stage. We validated its use when manipulating mechanistic models for scenarios where the primary focus is on estimating an average trend rather than the extremes of the distributions (RI-16-3; CI-12-2; CN-12-2). In this paper, we demonstrate improvements in accuracy and efficiency of parameter estimations and predictions in the context of HIV treatment when using SBA with results close to those obtained by analyzing the entire data at once. This type of method is particularly useful in scenarios that require adaptive decision-making (see Section 3.2 in the application for predicting the optimal treatment dose in PLHIV), or where data arrives incrementally over time (see Section 4.3 in the application of predicting COVID-19 epidemics).

The choice of the components (mathematical, statistical and observation models see Section 2.2) of a mechanistic model is a challenging process which requires confirmed expertise, advanced statistical methods, and the use of sophisticated software tools. The **procedure for constructing a model** is usually iterative: one adjusts a first model to the data, and diagnosis plots and statistical tests allow to detect possible misspecifications in the proposed model. A new model must then be proposed to correct these defects and improve the predictive abilities of the model. When building the statistical model, most of the common approaches consist in stepwise procedures consisting in testing the addition of variables forward and their elimination backward alternatively and progressing through the choice of models using an information criterion. A widely used approach is Stepwise Covariate Modeling (SCM) [98] which consists in an exhaustive search in the covariate space. Each covariate addition or deletion is tested in turn selecting models at each step leading to the best adjustment according to the objective criterion. Approaches such as Wald Approximation Method (WAM) [110] and COnditional Sampling use for Stepwise Approach based on Correlation tests (COSSAC) [13] are less computationally intensive as they use, respectively, a likelihood ratio test and a correlation test to move in the covariates space, which allows the evaluation of less models. All these methods are nevertheless computationally intensive as they require to re-estimate the model parameters and the likelihood many times. We proposed an **automated model-building** process by iteratively improving model components through a stochastic approximation approach (RI-22-1; HCI-22-3; HCI*-21-2). This algorithm called SAMBA (Stochastic Approximation for Model Building Algorithm) has been released in a R package *Rsmlx* on which I modestly contributed (Soft-23-1). See Section 6.3 for the full version of the article presenting this work. This method bears similarity to the Generalized Additive Model (GAM) [81, 128] but differs in its approach to parameter estimation. Instead of relying on Empirical Bayes Estimates (EBE), which are prone to shrinkage [144], this method utilizes the conditional posterior distribution of parameters. It uses a sample of the posterior parameters to build fast linear models linking them with covariates and proposing the most likely relationship. And, extension of these ideas opens the perspective for future research in model building strategies bridging the gap

to high dimension (see Section 2.6).

2.6 Perspectives

Below is a collection of perspectives, both short-term and long-term, that we aim to address by developing new methods for mechanistic models. Ideas marked with an asterisk (*) are considered major in terms of challenges and potential impact.

Advances in identifiability and optimal design: With recent advances, identifying structural identifiability of the mathematical model is no longer a major issue [121, 99, 87], and [6] for a review. However, it remains an open research question to carefully extend these concepts to the whole mechanistic model, including random effects [95]. In NLME models, ideal experimental conditions also include having data from an infinite number of subjects. However, the best strategy could rely in an optimal design in which a finite number of individuals are observed at a finite disjoint timepoints. The Cramer-Rao bound establishes a formal mathematical inequality stating that the covariance of any unbiased estimator is at least as large as the inverse of the Fisher Information Matrix (FIM). Thus optimizing a design is maximizing the FIM. The D-optimality criterion is widely used. It consists in maximizing the determinant of the FIM normalized by the number of parameters to be estimated. In NLME, the FIM can be computed numerically or approximated by a close-form solution [134]. Multiple algorithms have been proposed, implemented and compared for this purpose [146, 57]. An important practical path of advancing the research that we have in our SISTM group is to be able to use these methods more frequently in practice when designing new trials, based on the hypothesis that mechanisms of virus and immune responses are available and trusted for design (see Section 1.3.1 for the spirit). However, it may happen that some covariates are taken as explanatory covariates in the model and may be time-varying. This is for example the case for correlates of protection against an infection that are themselves likely to vary over time and impact the infectivity of a virus (see Section 3.4). This is also the case in COVID-19 transmission while changing the public health policies (see Section 4.3.1). Changing a parameter from constant to time-varying (because impacted by covariates) naturally influences structural identifiability and observability. Having a parameter that is time-varying can even improve the model identifiability [130]. Optimal sampling for these covariates, which are not by themselves the process of interest, is an open question, yet also to be extended in NLME settings. Finally, optimal design approach assumes that the true model underlying the data generation is known, which is often not the case in practice [3]. Thus, in term of methodology, one direction could be to extend the optimal design criteria to account for this model uncertainty, similar to model averaging [63]. This methodological aspect should be connected with the identification of suitable models for specific questions, as discussed in Section 3.5.

NLME Model Building: I am optimistic that this line of research will not only continue through the publication of independent academic studies but also through collaboration with the industry, notably with Simulations+ company, which is commercializing the Monolix Suite software.

- **Advances in structural model misspecification:** First, we believe it is important to detect when a model is not adequate for the data at hand. The optimal control approach has the advantage to provide an extra parameter - the control term or perturbation u. It is a function or set of variables that are manipulated to steer an ODE towards data. In our previous research, we did not fully exploit the optimal control perspective for misspecification quantification. Earlier studies have investigated the estimation of a perturbation term at the derivative level in a single-subject context using non-parametric methods to detect model errors [88, 60]. These methods, compared to those based only on data fitting criteria, tend to provide more sensitive statistical tests and can identify misspecification even in unobserved state variables. Our control-based approach could expand these tests to a population framework. For instance, within a Bayesian framework following [41], we could set a prior distribution for the controls and then assess it against the posterior obtained after inference. This work is meant to be done in collaboration with Quentin Clairon (Inria Sistm). Furthermore, when a large number of markers are available, it may be beneficial to identify which ones provide meaningful information to the model. This selection could be based on comparisons of restricted log-likelihood, although the statistical properties of this approach will require thorough investigations. An alternative could rely on using latent processes [188]. In this approach, the model would combine a multivariate linear mixed model and an ODE to model trajectories and temporal relationships between latent processes. In any case, pursuing this line of research presents a significant challenge in carefully and successfully utilizing the data available in a study.

- **Building variability models** Questions may arise regarding the strategy for **building variability models** within mechanistic models, i.e. which parameters should have a between-individual variability modeled by random effect. A traditional method involves adopting a SCM algorithm. However, this requires the estimation of model parameters at every step. We propose a method to reduce computation time by initiating only a few iterations of the SAEM algorithm from the previous stopping points, while allowing one of the random effects to tend to zero. This method facilitates rapid evaluation of a proxy for the complete likelihood, thereby swiftly guiding which random effects are the most likely to be added or removed. This approach has been presented at a French conference (CI-23-1) and is intended for collaboration with Marc Lavielle (Inria Xpop, yet retiring).

- **Extension to high-dimensional explanatory covariates:** When dealing with **high-dimensional covariates**, classical regression tools may overfit the data. A penalized approach, such as Lasso (Least Absolute Shrinkage and Selection Operator [195]), offers a solution by performing variable selection through the shrinkage of coefficients to zero, a feature not available in standard regression. Furthermore, Lasso aids in managing model complexity and preventing overfitting. Other penalization techniques, such as SCAD or MCP, and methods like ridge regression or elastic net, may also be relevant [147]. We have begun to extend covariate model building to high dimensions using these techniques. This effort, also presented at a French conference (CN-23-1), is conducted in collaboration with Auriane Gabaut, one of my Ph.D. students. Other methods developed in the team such as random forest [71, 24] should be also investigated.

- **Others:** Finally, questions regarding the order in which components (structural model, variability model, covariates model, correlations model, error model, observation model) should be built remain open, with no definitive answers applicable to all scenarios. Nonetheless, establishing guidelines would be beneficial for practitioners.

Toward high-dimensional mechanistic models (*): Once designed, practical non-identifiabilities may still exist in models and can be effectively detected using the profile likelihood method. To resolve these non-identifiabilities and achieve model identifiability, it's either necessary to simplify the model's complexity or to incorporate additional data. Techniques like profile likelihood-based model reduction [127] are crucial in this context, providing effective strategies for model simplification and data augmentation to improve model identifiability and reliability. Although tools for identifiability analysis can be improved, if data are not sufficient there will be no solution for analysis. Thus, we want to feed the model with all data generated in a multi-scale and multi-study approach. This includes feeding the model with fixed parameters that will be learned on previous studies. In particular, we think about pre-clinical studies (see Section 3.5 for details on future research on bridging between animal species). However, a particular idea in mind that I want to describe in this section is the deconvolution of gene expression data in the whole blood. Measurement of blood biomarkers, including transcriptomics, can be modeled using a set of ODE re-transcripting the temporal variations of the biomarkers and their inter-related trajectories. It could be seen as coupling a gene regulatory network [209] with more standard humoral dynamics models as developed in Section 3. The general framework can be written as:

$$\dot{X}(t) = F(t, X(t), \theta),$$

where $t \in [t_0, T] (0 \leq t_0 \leq T < \infty)$ is time, $X(t) = (B_1(t), \ldots, B_{p_B}(t), G_1(t), \ldots, G_{p_G}(t))^T$ is a vector

representing the p_B blood biomarkers and the gene expression level of gene $1, \ldots, p_G$ at time t, and $\dot{X}(t)$ is the first-order derivative of $X(t)$. F serves as the link function that quantifies the regulatory effects of regulator genes on the expression change of a target gene, which depends on a vector of parameters θ. In general, F can take any linear or nonlinear functional forms. It is important to note that, in this theoretical model, the ODE are high-dimensional. Specifically, p_B is on the order of magnitude of 10 or less, whereas p_G can be up to more than 10,000. With the current advancements in the estimation of NLME models, solving inverse problems in high dimensions becomes impractical. When dealing with high-dimensional systems, it may be practical to perform parameter estimation sequentially. For example, estimate a subset of parameters or a subset of ODE at a time, and iterate through the components of the system. In the specific application we have in mind, we want to use high-dimensional transcriptomic data to infer dynamics of cell populations. We postulate that data obtained from transcriptomics can be used with deconvolution methods to infer the composition or population of cells in a mixed sample [27, 12, 7]. Then mechanistic models can be fitted to these types of data. Deconvolution methods offer a solution to reduce the dimensionality of the vector $(G_1(t), \ldots, G_{p_G}(t))^T$. This reduction serves as a preliminary step, yielding a new set of new markers $DG_1(t), \ldots, DG_{p_{DG}}(t))^T$, where p_{DG} is approximately an order of magnitude of 10 or less. Consequently, the dimensionality of the new system, $p_B + p_{DG} < 20$, becomes manageable. We started working on a two-step approach - reduction of dimension based on deconvolution followed by mechanistic estimation - and demonstrated that transcriptomic data could help in identifying parameters for compartments of the ODE that are not observed by biological measurements (CI-23-2; CN*-23-1). A two-step approach capture the population aspect of the data, initially estimating a structural model for each patient, followed by an assessment of inter-patient variability factors. However, two-step approaches, which involves breaking down the analysis into two distinct phases, face a major disadvantage. Errors or biases in the first step can propagate into the second step, potentially leading to misleading results. In different yet comparable settings, our research demonstrates the superiority of a one-step mechanistic modeling approach, especially in terms of uncertainty propagation Sub-4. One-step approach for high-dimensional estimation, which simultaneously considers all aspects of the data and the relationships between them, is yet to be extended. This will be pursued using latent variable models in collaboration with Cécile Proust-Lima (Inserm Biostat). Latent variable models are statistical models that incorporate latent (unobserved) variables along with observed variables [159]. The challenge of constructing and inferring parameters in high-dimensional ODE is crucial for the future of mechanistic modeling and the advancement of new applicative approaches, as presented in Section 3.5.

Alternative estimation approaches: If there are advances in technologies or aforementioned deconvolution methods, our longitudinal biomarkers will transform into time series measured daily.

It opens the path to numerous new methodological developments applicable to our data.

- **Other Types of Differential Equations:** Firstly, there is potential for extension to more complex models such as Partial Differential Equations [32]. Indeed, spatial spread of virus is an important aspect in within-host modeling [78]. Secondly, Stochastic Differential Equations [52, 50] could be investigated, as they are commonly used to model cell growth dynamics [53]. Finally, to sustain the developments for multiscale between- and within-host modeling proposed in Section 4.4, we will develop new methods and explore already implemented approximate Bayesian computation for multiscale models of multicellular processes [2]. However, it has been shown that rich (as opposed to sparse) data is required to solve the inverse-problem in these types of models.

- **Sequential Data Assimilation:** As an example in the past, we did not manage yet to apply directly the population Kalman filter assimilation approach to fitting mechanistic data for virus or immune response dynamics due to sparse sampling of data. The original method for population Kalman filters (RI-22-5) proposed the explicit Euler method for numerical resolution of ODE. This method may fail when the ODE are stiff, which is the case in most of our applications. We initiated the implementation of alternative explicit methods such as the fourth-order Runge-Kutta, but found that explicit methods like Crank-Nicolson or BDF of order 2 tend to yield better results. It is important to note that, in contrast with explicit methods, implicit methods require executing Newton's algorithm one or more times per iteration. This generally leads to an increase in computational time making the problem whatsoever intractable. We intend to continue exploring modifications of the method with Annabelle Collin (Inria Monc) if data become highly time-continuous.

- **Curve Registration:** We would like to consider curve registration as a valid alternative to account for different timing of dynamics between individual for the mechanistic model. Curve registration, also known as temporal alignment, is a statistical technique used to align sets of curves or time series data so that they are in sync with each other in terms of certain features or landmarks. It assumes that the time has itself a individual-specific dynamic [163]. We have already initiated a project with Quentin Clairon from our team and a collaborator from Arizona State University, John Fricks. Results could become really valuable if data become more time-continuous.

- **Machine Learning Approaches (ML):** Finally, the advent of ML methods in the analysis of longitudinal repeated data in the recent years marks a significant turn in biostatistics. These methods excel in uncovering intricate patterns and relationships, even in large datasets. However, despite their sophistication, ML methods often face problems when the data points are not

evenly spread out over time or when some data is missing (see Section 4.4) or noisy, which is present in all our datasets. We will investigate the use of Reservoir Computing [124] as implemented in the Python library ReservoirPy [198] to perform forecasting task on viral dynamics and compare it with mechanistic models. We will also explore other approaches recently developed for between-host models [161] using long-short term memory recurrent network [72]. Given the sparse noisy nature of the data, we plan to perform various preprocessing steps such as multiple imputation and smoothing with a moving average to fill in the gaps. We will also propose an hybrid forecasting model in which the mechanistic model serves to provide complete de-noised smooth data, that are then assimilated by the ML algorithm. This work is meant to be done in collaboration with Xavier Hinault (Inria Mnemosyne) and Cécile Proust-lima (Inserm Bordeaux population health). This area of research might expand to include physically-informed neural networks, incorporating the concept of hybrid ODE [142]. Hybrid ODE are constructed by defining a loss function that includes terms for both the data discrepancy (how well the network predictions match the observed data) and the physical discrepancy (how well the network predictions comply with the governing mechanistic model). During training, the network learns to minimize this composite loss function, effectively fitting the data while also obeying the specified mechanistic model.

In conclusion, I am committed to continually updating the methods I use for solving inverse problems in NLME models and intend to persist in this endeavor.

Chapter 3

Within-host Modeling in Infectious Diseases

3.1 Bibliometry

Moving into a more applied realm, the second research direction introduces computational approaches to biology and medicine, specifically focusing on modeling the efficacy and impact of antivirals and vaccines. This section builds upon the theoretical methods developed in the first research direction, extending their application to practical scenarios. It encompasses not just the direct use of these methods, but also elaborates on a methodological framework designed for prediction and optimization of drug delivery. The manifestation of a viral infection in individuals varies and is host-dependent. Various techniques detect viral infection in biological samples. Quantitative methods are used to measure viral load or the concentration of specific proteins. For instance, the ELISA method detects certain antigens/antibodies, indicating the presence of a virus in a sample. Polymerase Chain Reaction (PCR) methods can detect viral genetic materials (DNA or RNA, with an additional transcription reverse step) even at very low concentrations. Functional assays of neutralization for antibodies are designed to measure the ability of these antibodies to prevent viral infections. Flow cytometry also allows to quantify the abundance of multiple types of cells such as B or T cells. This type of data will be modeled using **within-host mechanistic models** in this section. Table 3.1 showcases the scientific outputs in this field. **This theme gave birth to 16 articles in peer-reviewed international journals, 21 international conferences, 4 national conferences and 6 seminars. It is also funded by 2 international grants and 2 national grants.** Even though most models and methods could be easily extended to other infectious diseases, we particularly worked on HIV, Ebola virus, Nipah virus, and SARS-CoV-2. The interrelation of all these works is explained in the next subsections.

Peer-reviewed International Journals		**First**	**Second**	**Before Last or Last**	**Other**
Nature Communication					RI-21-3
Advanced Drug delivery Reviews		RI-13-2			
eLife			RI-22-4		
eBioMedicine					RI-22-3
Nature Vaccine			RI-23-7		
Journal of Clinical Immunology					RI-23-4
PLOS Computational Biology			RI-23-5		RI-14-1; RI-23-8
Journal of Virology					RI-19-1; RI-16-1
Journal of Theoritical Biology					RI-20-1
Statistical Methods in Medical Research				RI-21-2	
Biometrics		RI-12-2			
Statistics in Medicine				RI-18-2	
Annals of Applied Statitistics					RI-17-3
Total (16)		2	3	2	9
International conferences	**Invited**	**Speaker**	**Mentee**	**Poster**	**Other**
CROI				HCI-23-4; HCI-20-1; HCI-18-2; HCI-18-1	
World congress B&C Pharmacology	HCI*-23-3				
PAGE			HCI-23-2; HCI-22-2; HCI-21-4; HCI-18-3	CI-19-2	
Society Mathematical Biology	HCI*-23-1				
Canadian Applied Mathematics Society	HCI*-22-1				
CMStat			HCI*-21-2		
Workshop on Virus Dynamics	HCI*-19-2				CI-19-5
HIV Dynamics and evolution	HCI-13-1				
ISBC			CI-21-1; CI-20-1		
IBC			CI-18-2		CI-14-2
Other					CI-17-1
Total (22)	5	0	8	5	3
National conferences	**Invited**	**Speaker**	**Mentee**	**Poster**	**Other**
ANRS MIE AC	CN*-21-3		CN-22-2		
JdS SFdS			CN-21-2		
GDR Stat & Santé ou Mamovi		CN-17-3			
Total (4)	1	1	2	0	0
Seminars (6)	S-23-1; S-22-1; S-18-1; S-17-2; S-13-2; S-13-1				
International Grants (2)	IG-19-2; IG-18-1				
National Grants (2)	NG-23-1; NG-18-1				
Sofwares (0)					

Table 3.1 – Bibliometry for axis "Within-host modeling in infectious diseases"

3.2 Virus Dynamics Models

3.2.1 Generalities on Modeling HIV Dynamics

The new field of viral dynamics, based on within-host modeling of viral infections, began with models of human immunodeficiency virus (HIV), but now includes many viral infections. Early research focused on viral dynamics after the initiation of treatment. This work began with the biological observation that the virus V is produced by infected cells at a rate denoted by β and dies at a constant rate δ_V. This concept was expressed through the differential equation $dV/dt = \beta - \delta_V V$. This basic analysis demonstrated a rapid turnover of HIV viruses [205]. The model was later extended to account for CD4+ T cells (CD4 in short) [85]. In the model, uninfected target cells T are produced at a constant rate λ, die at a rate δ_T per cell, and get infected by the free virus at a rate described by the mass action term βVT. This infection process leads to the creation of productively infected cells I, which die at a rate δ_I, higher than δ_T, reflecting the viral impact on reducing the lifespan of infected cells. Finally, free viruses are produced by infected cells at a rate p per cell and are cleared from circulation at a rate δ_V per virus. This is transcripted in Equation 3.1 and Figure 3.1.

$$\begin{cases} \frac{dT}{dt} &= \lambda - \delta_T T - \beta VT \\ \frac{dI}{dt} &= \beta VT - \delta_I I \\ \frac{dV}{dt} &= pI - \delta_V V \end{cases} \tag{3.1}$$

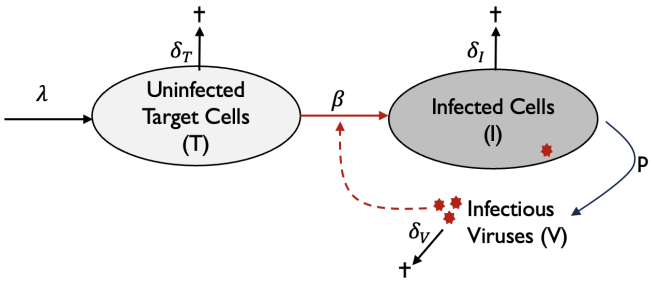


Figure 3.1 – Target Limited virus dynamics model

Several reviews discuss variations of this fundamental HIV dynamics model [151, 150]. These models can be specifically tailored to the particular mechanism of interest or research question. Extensions of the model may include more complex CD4 maturation processes [55], the consideration of multiple viral quasi-species (both infectious and non-infectious) [145], the inclusion of effector cells [15], CD8 modelling [170] or the dynamics of HIV latency in reservoirs [169]. The choice and complexity of the model are influenced by the available data to ensure identifiability without needing to fix most of the model's parameters.

Because theoretical research from Section 2.2 demonstrates the feasibility of utilizing this mechanistic model platform for **predicting counterfactual scenarios**, we aim at using mechanistic model to investigate the change of dose and the timing of dose administration in a personalized medicine framework. It primarily depends on updating population parameters from a mechanistic model using empirical data for a new subject. This is most of the time achieved through Empirical Bayes Estimates (EBE [139]). This process enables the refinement of the subject's random effects, facilitating predictions about its response to new interventions. These interventions effects may have been learned from a different set of patients.

3.2.2 Generalities on Personalized Medicine

Recent systematic review of the literature indicates that prediction under hypothetical intervention for **treatment optimization** is still an emerging topic with most work appeared later than 2015 [122]. **Optimal treatment regimes or treatment drug monitoring** involve sequential decisions [119]. In these trials, a treatment regime is a sequence of decision rules, each rule corresponding to a decision point. These rules map the accumulated patient information to a recommended intervention. The approach essentially tailors the treatment plan based on the evolving health status of the patient, allowing for a more personalized and effective treatment strategy. There are currently at least three distinct types of optimal treatment regimes : regression-based mainly with Q- and A- learning (see [30] and references therein), direct search which construct an estimator of the mean cumulative utility (see [216] and references therein) and model-based. In our work we focus on the later for which G-computation is an option [215] as well as Bayesian approach [171, 79, 211] (see later in Section 3.1 for our developments) and learning methods (see [73] and references therein as well as Section 4.3 for our developments). At the time of this research (2013), **model-based strategies in clinical decision-making** were less mainstream than they are today. We authored a statement paper on how mechanistic models can inform personalized medicine (RI-13-2; S-13-2). In a model-based approach, the idea is to learn from the data collected from an individual to forecast the best choices likely to produce the best clinical outcomes. It has been shown possible in many pharmacometrics models using clinical trials simulations, for example in herpes viruses [175]. Control theory has been suggested to find the optimal interventions that reduce a given cost function [107]. Quadratic cost functions weighting system response (viral load or CD4 count) and side-effects of the drug have been proposed [185]. Although conceptually interesting, this approach is not realistic because (i) the model is not known, (ii) the choice of the cost function is debatable, (iii) the treatment cannot be continuously adapted. More recent approaches tend to get free from cost functions [11] (as we do in the following) and do not aim at adapting the dose continuously [200].

In my work, **the effective reproductive number** (R_0 before any intervention and R_t after) plays

an important role. It represents the average number of secondary infections produced by a single infected individual in a completely susceptible host population. It is a key concept in epidemiology, indicating the contagiousness of an infectious disease, see Section 4.1. In the within-host scenario, it describes how efficiently a pathogen can replicate and spread within an individual host. A higher R_0 value suggests a more aggressive infection, potentially leading to faster disease progression or greater virulence. On the contrary, R_0 below one signifies that the infection is unlikely to sustain itself. Thus, in our work on optimization of delivery of treatments this represents the outcome we will try to control.

3.2.3 Bayesian Optimization of Delivery of ART

Antiretroviral therapies (ART) is a treatment regimen for HIV/AIDS that uses a combination of several antiretroviral drugs. The aim of ART is to reduce the amount of HIV viral load to undetectable levels. This approach helps in improving the immune functions and slowing the progression of HIV to AIDS. ART has significantly improved the life expectancy for people living with HIV (PLHIV) to few years lower compared to people non-infected with HIV [197].

Our initial work was rather conceptual. It proposes **a personalized medicine approach for HIV treatment based on controlling reproductive number** (RI-12-2). The strategy aims to determine the minimal ART dose that keeps viral load undetectable (i.e. R_0 below one), thereby reducing side effects. The flowchart of the strategy is provided in Figure 3.2. For a new individual, the algorithm involves estimating individual parameters using patient data observed after the initiation of a standard dose. The estimation is done using a mechanistic model similar to Equation 3.1 with an EBE approach for updating individual parameters. Then, we sample the posterior distribution of R_0 for this specific patient using MCMC sampling. The optimal dose d_{opt} is defined as the dose for which the probability of having a R_0 greater than one is small, typically lower than 5%. Finally, we observe the dynamics of the patient under this new dose and readjust iteratively until convergence. We then extend this work based on a stochastic model for treatment optimization Sub-1. Later, we successfully predicted outcomes of ongoing trials on **Short-cycles treatment interruptions**, in which patients take their treatment only few consecutive days in the week (HCI-18-2; HCI-13-1; CN-17-3; S-13-1). We estimated the effect of various ART with mechanistic models on the HIV Aquitaine ANRS CO3 Cohort [193] and predicted the results of the Breather trial which investigated 5/7 designs (5 days on, 2 days off ART) in adolescents. They showed a sustainable non-inferiority of virological suppression compared to continuous ART [19]. We were able to predict this result. Finally, we encapsulated all the steps of these methods to develop an **in silico trial pipeline** (S-18-1; S-17-2). See Section 3.5 for details on elaborating a pipeline that will provide a **digital twin** for drug and vaccine development in infectious diseases.

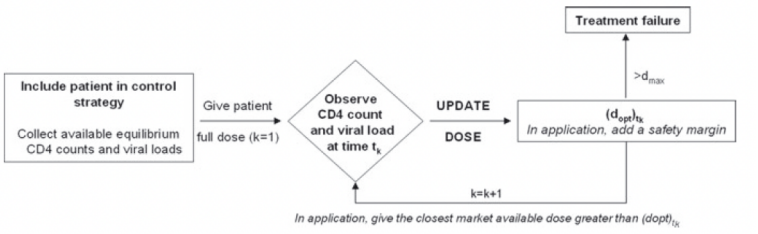


Figure 3.2 – Flow chart for the control strategy procedure: individualized dose monitoring for ART dose in PLHIV (Figure from RI-12-2).

3.2.4 Modeling T cells proliferation and optimization of timing delivery

We developed a comprehensive mathematical analysis of the effects of **Interleukin-7** (IL-7) on CD4 restoration in PLHIV. We model CD4 proliferation using mechanistic model with two compartments of proliferating (P) and quiescent (Q) cells. Of note, other modeling approach such as size-structured model for cell division could have been proposed [54]. Q cells are created at rate λ and start proliferating at rate π. Each P cells results in the formation of 2 new Q cells at a rate ρ. Cells die at rates μ_P and μ_Q. This is transcripted in Equation 3.2 and Figure 3.3.

$$\begin{cases} \frac{dQ}{dt} &= \lambda + 2\rho P - \pi Q - \mu_Q Q \\ \frac{dP}{dt} &= \pi Q - \rho P - \mu_P P \end{cases} \tag{3.2}$$

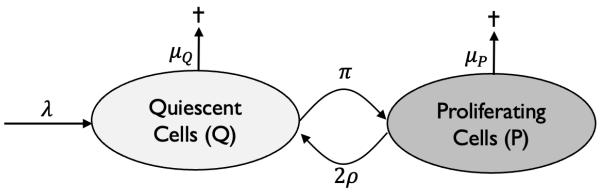


Figure 3.3 – CD4 Proliferation Model

The objective was to understand and quantify whether the observed changes in CD4 counts could be attributed solely to increased peripheral proliferation or if other mechanisms, such as improved cell survival or thymic production, also play significant roles (RI-14-1; CI-14-2; CI-17-1). In particular, we demonstrated that a decrease of the loss rate of the quiescent CD4 (μ_Q) is the most probable explanation. This platform served to explore the potential of repeated IL-7 cycles for sustained CD4 restoration in PLHIV (RI-17-3; HCI-18-3). These results unlocked the design of new clinical trials for repeated IL-7 injections to sustain high level of CD4 in PLHIV (Inspire trials). Finally, we designed **adaptive protocols based on Bayesian predictions** from a mechanistic model of the effect of IL-7 on CD4 counts (RI-18-2; CI-18-2). Figure 3.4 presents two adaptive treatment protocol we proposed. One based on the fixed visit times and adaptive criterion of injection of IL-7 (AC) and one based on

individualized visit times (AT). We show that AT slightly reduce the number of visits compared to AC, while keeping other criteria such as CD4 levels unchanged. In comparison to classical approaches rooted in optimal control [148], our method offers distinct advantages. Firstly, it operates under the assumption that parameters are not pre-defined, allowing for adaptation of IL-7 injections as estimations of individual parameters and predictions are refined with new data in time. This dynamic approach aligns with the concept of dynamic drug monitoring as described in [141]. Secondly, our statistical treatment optimization approach is less computationally intensive, as it does not involve searching for an optimal strategy across the entire space of potential strategies. **This project really illustrated how new trials can be guided by simulations and extrapolations from mechanistic models.** See Section 3.5 for details on ways to explore new approach for treatment design.

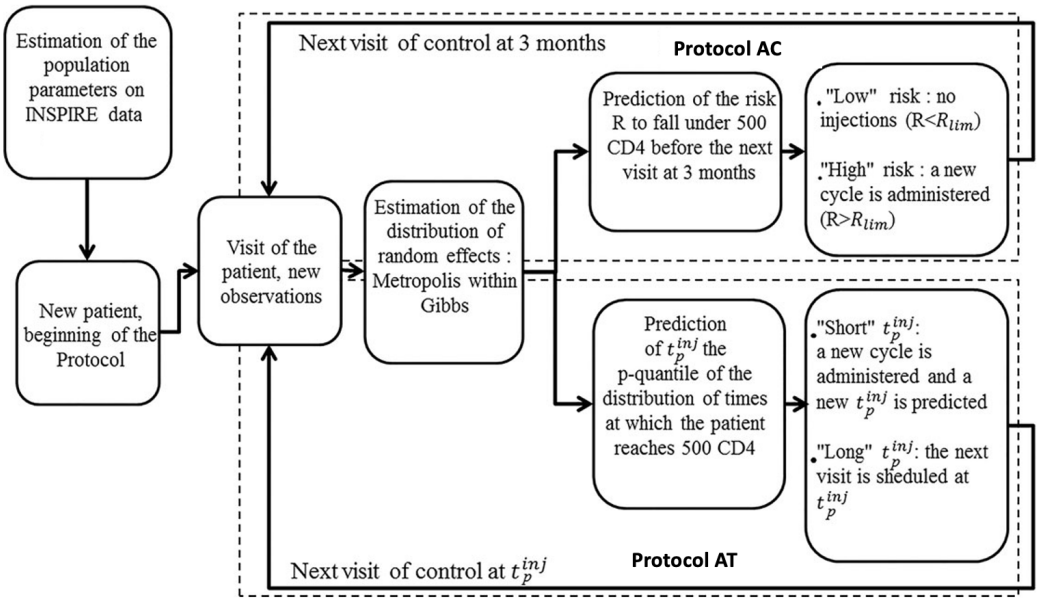


Figure 3.4 – Flowchart illustrating two protocols. Adaptive criterion of injection (AC) protocol: the visits are every 3 months and the decision to administer a new cycle is based on the predicted risk R to fall under 500 CD4 count before the next visit. Adaptive time of injection (AT) protocol: the times of visit are predicted based on the time at which the patient is supposed to reach the 500 CD4 count limit, and a new cycle is administered if this predicted time is too short (Figure from RI-18-2).

3.2.5 Methods to evaluate vaccine efficacy

There have been a long research on vaccine against HIV. Although we got involved in some work regarding pre-clinical development of prophylactyc vaccines (RI-16-1, in which we ran a basic survival analysis), we are particularly interested in the development of **therapeutic vaccines for PLHIV**, also referred as HIV cures. Therapeutic vaccines for HIV are designed to enhance or induce immunity in order to alter the disease's course. Despite the success of ART in managing HIV, these therapies alone

are unlikely to eliminate the HIV reservoir and cure the infection. **Analytical treatment interruption** (ATI) in HIV management involves temporarily stopping ART. ATI is used in research to evaluate the efficacy of HIV therapeutic vaccines. A consensus paper on the general rules to run ATI trials has been published elsewhere [100].

Viral load is most of the time of primary interest in ATI trials. First, this marker is always left-censored due to detection limit of biological measurements, i.e. the exact viral load is unknown and is effectively "censored" at a lower limit. Second, several recommendations have been proposed to limit potential risks, in terms of morbidity, mortality, disease progression, HIV transmission, emergence of new drug-resistance, development of neurological or cardiovascular disorders [101]. In particular, regulatory guidelines mandate ART resumption for patient safety when the viral load reaches a certain threshold. This is similar to informative loss of follow-up which is a type of **informative missing data**. Thus, end-of-the-study viral load (setpoint) value in itself cannot be primary outcome of these trials. In our work, we statistically compared endpoints as primary outcome for trials (HCI-23-4; HCI*-19-2; Sub-5). We investigated among others time-related criteria (slope of rebound, time to rebound, to peak viral load, to setpoint, and to specific viral load thresholds) and viral load magnitude-related criteria (peak viral load, setpoint and time-averaged area under the curve nAUC). All these criteria are depicted in Figure 3.5. We showed that **time-averaged Area Under the Curve** (nAUC) is the most robust indicator. Subsequently, with my PhD student Marie Alexandre, we proposes a statistical testing of the nAUC strategy (RI-21-2; HCI*-21-2; CI-20-1; CN-21-2). See Section 6.4 for the full version of the article presenting this work. The viral rebound trajectories which are partially observed are approximated using spline-based mixed-model. Then, time-averaged AUC are derived from the splines regression coefficients to perform mean-difference t-test between groups. See Section 3.5 for details on extension of this method using mechanistic models and manner to account for model uncertainty using model averaging.

Mechanistic modeling HIV rebound after ATI using dynamic models presents several challenges. A major challenge is the latent HIV reservoir, where the virus remains hidden and inactive within cells, evading the immune response and ART. This reservoir can persist despite long-term therapy, leading to viral rebound when treatment is interrupted. However, the time to rebound is individual-specific and stochastic as a rare event process. Together with collaborators (Alison Hill - Harvard University; John Hopkins Hospital), we developed a good model for rebound in HIV accounting for the stochasticity of the phenomenon (HCI-18-1; CI-19-2). The model was specifically developed to be flexible enough to capture rebound kinetics both in the regime where latent cells reactivate frequently and rebound occurs rapidly, and in the regime where **reactivation from latency** is rare and there are stochastic delays until the first fated-to-establish lineage exits the reservoir [84]. Briefly, free viruses V enter target cells T (with infection rate β), producing infected cells I. Infected cells in

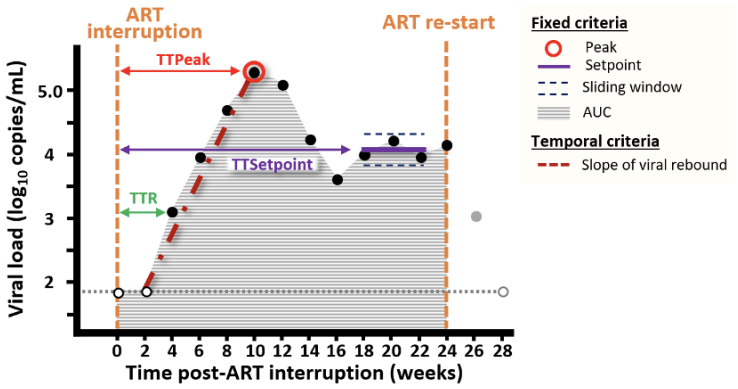


Figure 3.5 – Description of virological endpoints. Black and white dots represent detectable and undetectable viral load measurements collected during the ATI phase, respectively, while gray dots represent measurements taken after ART re-initiation. The horizontal gray dotted line represents the detection limit. Abbreviations: ART, antiretroviral treatment; TTR, time to rebound; TTsetpoint, time to setpoint; TTPeak, time to VL peak; AUC, area under the curve; VL, viral load (Figure from Sub-5).

turn release free virus (rate k). Long-lived precursor immune cells P which encounter viral antigen proliferate ($p(V) = pV/(V + N_P)$) and produce short-lived effector immune cells E. Effector immune cells eliminate some infected cells before they contribute to ongoing infection by producing new virus ($\beta(E) = \beta/(1 + E/N_E)$). A fraction f of expanded precursor cells revert to the precursor state after encountering antigen, forming immunological memory. Both uninfected target cells and precursor immune cells are produced at a constant rate (λ and m, respectively). While during acute infection m likely represents activation of naive cells, during viral rebound, it may be dominated by reactivation of memory cells. Latently infected cells reactivate with rate a (or equivalently, every t_a days on average) to become productively infected cells. Virus is cleared at a rate c and each cell type i dies with death rate d_i.This is transcripted in Equation 3.3 and Figure 3.6.

$$\begin{cases} \frac{dT}{dt} &= \lambda - \beta TV - d_T T \\ \frac{dI}{dt} &= a + \frac{\beta TV}{1+(E/N_E)} - d_I I \\ \frac{dV}{dt} &= kI - cV \\ \frac{dP}{dt} &= m + p(1-f)\frac{V}{V+N_P}P - d_P P \\ \frac{dE}{dt} &= pf\frac{V}{V+N_P}P - d_E E \end{cases} \tag{3.3}$$

The study includes data from **non-human primates (NHP)** trials involving immunotherapies like TLR7-agonist, therapeutic vaccines, and monoclonal antibodies (Sub-2; HCI-20-1; CI-19-5). We showed a modest effect of these strategies in NHP. However, it has been demonstrated in [164] that treatment strategies providing modest but continuous improvements in reservoir clearance rates lead to quicker cures than abrupt, one-time reductions in reservoir size. Furthermore, the role of CD4 cell

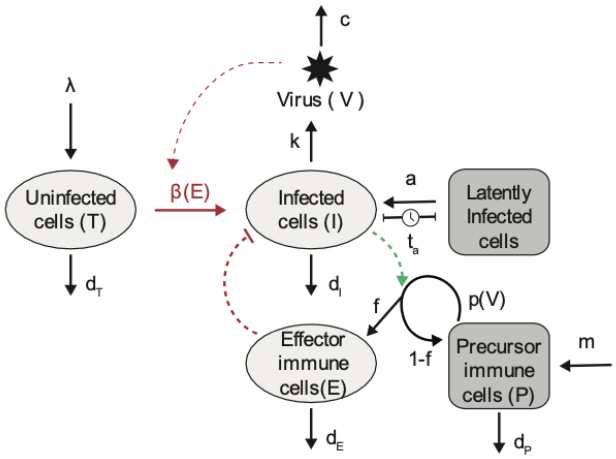


Figure 3.6 – Schematic of the viral dynamics model with latent infection and an antigen-dependent immune response (Figure from Sub-2).

turnover in maintaining HIV persistence during antiretroviral therapy remains to be fully elucidated [14]. Our preprint (Sub-2) also aims to predict the outcomes of these immunotherapies in human clinical trials. This marks a significant initial step in my research in connecting findings from NHP studies to human simulations through the use of mechanistic models. And, extension of these ideas of cross-species mechanistic predictions opens the perspective for future research (see Section 3.5).

3.3 Immune Dynamics Models

3.3.1 Generalities on modeling Humoral Response

The immune system comprises a vast array of diverse cells that interact with each other and the rest of the body to protect against external entities. It consists of two main parts: the innate immune system, which includes generalist cells and components forming the first line of defense against external agents, and the acquired immune system, activated by **T and B lymphocytes** in response to a pathogen, providing specific defense. Unlike the constant presence of innate immunity actors, acquired immunity cells are less abundant before the first infection, and developing an acquired response can take days, during which time the virus can proliferate. The activation of the acquired immune system leads to the recruitment of specific memory lymphocytes, offering long-term protection against infections by the same pathogen. This immunity is also marked by the presence of antibodies in the blood, specifically targeting viral proteins or nucleic acids.

Although mechanistic models can describe these processes, the literature is still very sparse. A recent review of the literature identified only 8 ODE-based modeling approaches of humoral immunity [70]. One of the most exhaustive model allowing to account for establishment and maintenant of

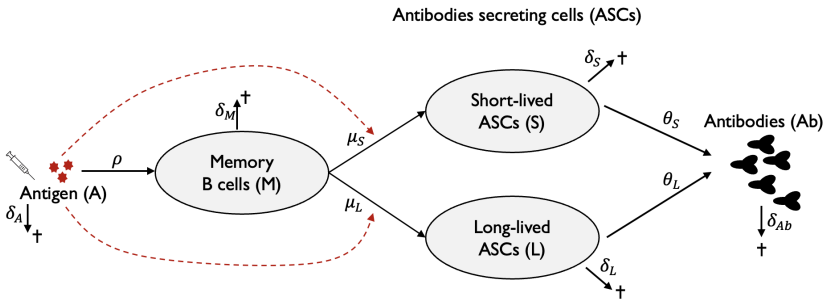


Figure 3.7 – Schematic representation of humoral response Model.

the immune vaccinal responses has been developed with collaborators (RI-20-1; CN*-21-3). To understand the **dynamics of antibodies**, it is possible to developed a model incorporating a population of **antibody-secreting cells (ASC)**. However, to explain the long-term persistence of the antibody response, it is needed to extend it to include two distinct ASC populations [68], which consists in **short-lived and long-lived populations**. This concept was further refined using ODE [208, 5]. Building on these ideas, we introduced a **memory B cell** compartment into our model, akin to the approach [42], where memory B cells can differentiate into ASC. This integration of processes resulted in Equation 3.4 and Figure 3.7, forming our mechanistic model.

$$\begin{cases} \frac{dA}{dt} &= -\delta_A A \\ \frac{dM}{dt} &= \rho A - (\mu_S + \mu_L) AM - \delta_M M \\ \frac{dS}{dt} &= \mu_S AM - \delta_S S \\ \frac{dL}{dt} &= \mu_L AM - \delta_L L \\ \frac{dAb}{dt} &= \theta_S S + \theta_L L - \delta_{Ab} Ab \end{cases} \tag{3.4}$$

The model considers three B cell populations: memory B cells (M), short-lived ASC (S), and long-lived ASC (L), along with **antigen** concentration (A) introduced through immunizations, and antibody concentration (Ab). The reaction in the model begins with the detection of an antigen A (which can be presented by vaccination or natural infection), leading to the formation of M cells at a rate denoted by ρ. These M cells differentiate into short-lived (S) and long-lived (L) ASC at rates described by the mass action terms $\mu_S AM$ and $\mu_L AM$. These ASC then produce Ab at a rate θ. All components X in the model decay at a rate represented by δ_X. While this model is an effective theoretical tool for conceptualizing and understanding the humoral response, it has a large number of parameters. This complexity makes the model structurally unidentifiable when only observing antibody levels. One potential solution is to design studies where proxies for M, S, and L cells are measured. However, currently, such measurements are not readily available. In the future, it might

be possible to achieve this using transcriptomics, as discussed in Section 2.6. Because this approach is not mature to be used yet, to effectively use this model, it's necessary to either fix many parameters through sensitivity analysis or develop simpler, rescaled models. The subsequent sections propose ways to address these challenges.

3.3.2 Maintenance of vaccinal response against Ebola

Having an effective vaccine against the **Ebola virus** hemorrhagic fever has become a public health imperative since the 2014 epidemic in West Africa. The **EBOVAC series grants** - including EBOVAC 3 for which I was involved (IG-18-1) - has contributed to meeting this challenge by providing evidence of the safety and immunogenicity of a vaccine developed by Janssen Pharmaceuticals. Data for these trials have been published in a serie of papers [181, 156, 16, 93]. The first goal was to understand the long-term immune response to the vaccine. Using a simplification of the mechanistic model from Equation 3.4 with only (S, L, Ab), we modeled the antibodies decline from the peak 7 days after last vaccination to last follow-up. The simpler model writes in Equation 3.5 and Figure 3.8, with $\phi_S = \theta_S S_0$ and $\phi_L = \theta_L L_0$, where S_0 and L_0 are the initial conditions at equilibrium of memory cells 7 days post last vaccination. Since both short-lived and long-lived ASC populations are unobservable, the parameters (θ_S and θ_L) and initial conditions cannot be identified solely based on observations of antibodies.

$$\frac{dAb}{dt} = \phi_S e^{-\delta_S t} + \phi_L e^{-\delta_L t} - \delta_{Ab} Ab \tag{3.5}$$

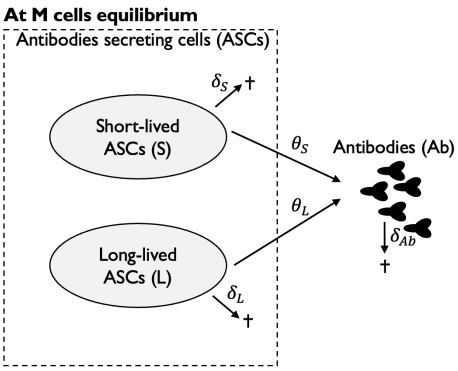


Figure 3.8 – Model for persistence of humoral response.

Multiple regimens that were tested in phase I (including different vectors order and timing - studies EBL1001, EBL1003, EBL1004) on 44 participants were evaluated and our modeling helped selecting Ad26.ZEBOV/MVA-BN-Filo with a booster dose at day 57 and predicting its effect (RI-19-1). This first piece of work predicted the durability of the response to be sustained more than 7 years. Using follow-up data from phase II/III (EBL2001, EBL2002, EBL3001) on 443 participants, we showed

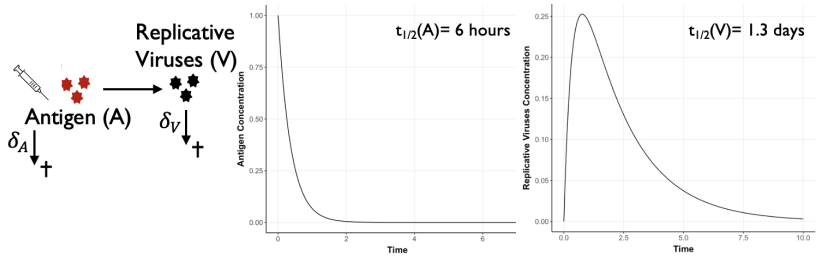


Figure 3.9 – Schematic of antigen presentation in replicant vaccine viral vector.

that immunity is likely to be sustained for more than 15 years using this strategy (RI-23-7; HCI-21-4). Using model building strategies defined in Section 2.5, we also investigated the effect of covariates on the humoral response dynamics. The humoral response's predictors include geographic region, where Europeans demonstrated longer antibody persistence compared to Africans; sex, with women showing longer antibody persistence; and age, where younger participants exhibited higher antibody production. These conclusions are yet to be extended in larger studies.

In the **Prevac/Prevac-Up trial** (funded by IG-19-2 for which I am workpackage leader), more than 3000 individuals were followed-up after Ebola vaccination using two strategies Ad26.ZEBOV/MVA-BN-Filo by Janssen Pharmaceuticals or rVSV-ZEBOV-GP by Merck. Data have been described in [192]. The idea is to understand if it exists important predictors of vaccine response such as demographic covariates. We built a linear mixed-effects model to start to answer this question (Sub-3), and mechanistic development are yet to be extended. One of the important mechanistic differences between the two vaccine platforms tested is that while Ad26 and MVA viral vectors are non-replicating, rVSV replicates within the host after participant immunization. This has been demonstrated by shedding studies conducted in both blood and saliva [165]. Consequently, the antigen presentation described in Equation 3.4 is likely to vary depending on the vaccination group. Explored features **accounting for replication**, akin to pharmacokinetic models, are presented in Equation 3.6 and Figure 3.9. However, due to sampling constraints, there is limited information available regarding this aspect of the dynamics. Nonetheless, we will aim to address the question of differing mechanisms of humoral dynamics between vaccine groups. The mechanism will also differ by individual. This would open the perspective of **optimization of vaccine delivery using a vaccinal digital twin**, see Section 3.5.

$$\begin{cases} \frac{dA}{dt} &= -\delta_A A \\ \frac{dV}{dt} &= A - \delta_V V \end{cases} \tag{3.6}$$

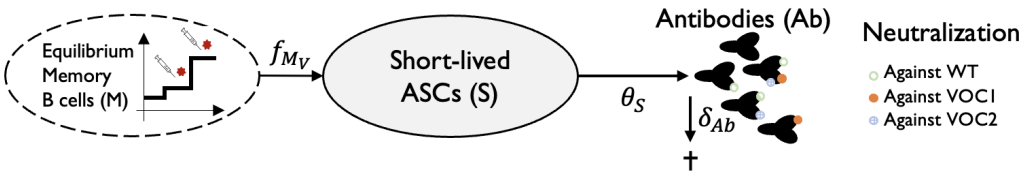


Figure 3.10 – Model for establishment of humoral response.

3.3.3 Establishment and maintenance of vaccinal response to SARS-CoV-2

The **COVID-19** pandemic, unprecedented in modern times, caused a global shift in research focus. Quickly, it became a priority for the scientific community, redirecting efforts from various fields towards COVID-19 research. We used our mechanistic tools to understand **humoral response and its neutralisation abilities** in humans. We rescaled and reduced the models developed for human response in Ebola from Equation 3.4 to be able to estimate it from antibodies data. Moreover, we demonstrated that a single compartment of ASC was sufficient to accurately model the dynamics, suggesting that there is no long-term response to the investigated vaccine. The new model is fundamentally based on the assumption that a steady state is rapidly achieved by M cells following vaccination using step function. These M cells mature into S cells at a rate of f_{M_V}, while the antigen, degrading at a rate of δ_V after each injection at time t_V ($V = 1, 2, 3$), is present. S cells die at a rate of δ_S. Antibodies Ab are then produced at a rate of θ_S and degrade at a rate of δ_{Ab}. Moreover, we jointly model the binding (Ab) and neutralizing (as measured by ED_{50} for each **variant of concern** (VoC) ν, which refers to the effective concentration of antibody required to neutralize 50% of the virus in an assay) by assuming that neutralizing antibodies are a proportional time-varying fraction of binding antibodies depending on the VoC (f_ν) and enhanced with new injections in a VoC-specific manner (f_V^ν with $V = 1, 2, 3$). This is described in Equation 3.7 and Figure 3.10. In other words, there is a decrease of neutralization against VoC compared to Wild-type strains (WT), an increased neutralization with the subsequent doses of vaccine, and the incremental increase in neutralization is more pronounced for the subsequent injections with Omicron VoC compared to WT.

$$\begin{cases} \frac{dS}{dt} &= f_{M_V} e^{-\delta_V (t - t_V)} - \delta_S S \\ \frac{dAb}{dt} &= \theta_S S - \delta_{Ab} Ab \\ ED_{50}^\nu(t) &= F(\nu, t) Ab(t) \\ F(\nu, t) &= f_\nu f_V^\nu(t) \end{cases} \tag{3.7}$$

See Section 6.5 for the full version of the article presenting this work. It allowed us to derive the **expected longevity of vaccine protection in humans** after **Pfizer Bnt162b2 vaccination** against each VoC up to the omicron era (RI-23-5; HCI-23-2; HCI*-22-1; CN-22-2; S-23-1; S-22-1). The mean

duration of detectable neutralizing capacity varied between VoC. For non-Omicron VoC, it ranged from 348 to 587 days, and for Omicron VoC, it was between 173 and 256 days post-third vaccination. The study highlighted the significant role of multiple vaccine doses in enhancing both the quantity and quality of the humoral response, particularly against emerging VoC. However, it should be considered that there is no definitive proof indicating what level of neutralization might be sufficient to protect against infection or severe illness. This is partially adresse by the work on **correlates of protection** (CoP) started in the next section 3.4. In our work, we did not address cross-immunity, although it is relevant [176]. Moreover, the protection responses are possibly multifactorial, including other markers such as T response [202]. On this note, in a side work using the French Covid cohort, we also investigated with basic descriptive statistics the impact of T cell response on protection against severe infection (RI-23-4). See Section 3.5 for details on exploration of animal platforms to investigate these questions.

3.4 Joint Modeling of Virus and Immune Dynamics Models

During the pandemic, we have been involved in research testing for a new vaccine platform against **SARS-CoV-2** (RI-21-3; RI-22-3). This vaccine platform targets the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein to CD40-expressing antigen-presenting cells, leveraging their immune-stimulant properties. The vaccine is tested on NHP, a platform beneficial for vaccine testing due to their close physiological and genetic similarities to humans. These platforms are vital in early vaccine development stages, particularly because it's possible to expose NHP to the pathogen to evaluate their dynamic response to infection. This method provides critical insights into the vaccine's potential effectiveness and safety before human trials. Using this unique data, we investigate the effect of VoC of SARS-CoV-2 on the viral dynamics as described in Equation 3.1 (RI-23-8). With collaborators from Inserm Paris IAME, Jérémie Guedj and students, we demonstrated that despite lower viral loads, Omicron maintained high levels of infectious particles over time, suggesting its increased transmissibility stems from prolonged infectiousness rather than higher viral load. A subsequent step would be to correlate these results with findings from extensive human screening data [59, 184], see Section 3.5 for bridging between species. It has been shown that while viral load is a reliable predictor for transmission [129], it remains inconclusive or not a strong marker for infection severity and disease progression [39]. Therefore, it's necessary to explore other markers of immune system dynamics for a more comprehensive understanding of these processes. A CoP is a specific immune response to a vaccine or infection that is statistically associated with protection against a disease. It serves as a surrogate marker indicating the presence and level of immunity. Identifying these CoP is crucial for evaluating vaccine effectiveness and guiding vaccine development [96]. Binding antibod-

ies to SARS-CoV-2 and in vitro neutralization of virus infection are clearly associated with protection [105, 214, 58, 64]. However, the respective contribution to virus control in vivo remains unclear, and many other immunological mechanisms may also be involved, including other antibody-mediated functions [135, 191], as well as T cell immunity [133]. Furthermore, CoP may vary between the vaccine platforms [154, 40].

In term of methods, various approaches has been proposed to identify CoP. NHP studies offer a unique opportunity to evaluate early markers of protective response [140, 62]. The concept of CoP is deeply rooted in the field of causal inference. A CoP is considered as such if knowing the effect of a vaccination on the CoP allows for the prediction of the vaccine's impact on clinical outcomes [97]. This aligns with the statistical definition of a surrogate marker, where the CoP acts as an intermediary predictor of the vaccine's efficacy in terms of real-world health outcomes. While some studies focus on patient-level data, examining correlations between markers and outcomes [36, 106], other use meta-analysis in finding CoP. This later approach involves pooling data from multiple studies to identify common immune responses that protect against disease, using regression techniques like linear and Cox models [75, 174]. Identifying CoP requires acknowledging that correlation does not imply causation, recognizing multi-dimensional mechanisms, and avoiding off-target effects (unintended consequences of an intervention) [67, 201]. Therefore, defining a CoP on the causal pathway is essential, aligning with the principles of mechanistic modeling, this defining a **mechanistic correlate of protection** (mCoP) [155]. We propose a method to define a mCoP using mechanistic models. This method validates whether a covariate serves as a surrogate marker in accordance with Prentice's definition [158]. See Figure 3.11 for a representation of CoP with Directed Acyclic graphs and Prentice's definition. Essentially, it allows for the evaluation of a covariate's role in mechanistically linking vaccination to clinical outcomes (RI-22-4; HCI*-23-3; HCI-22-2). See Section 6.6 for the full version of the article presenting this work. We demonstrated the crucial role of antibodies' neutralization function in the infection process, which is consistent with literature [105] and has been validated on two external studies. In brief, the method is as follows:

1. **Mechanistic modeling**: Fit a virus dynamics model to the data obtained after a challenge with the desired virus, without ($\mathcal{M}_0$) and with adjusting for intervention effect ($\mathcal{M}_G$).

2. **Correlation of the prospective mCoP with the outcome ($S \to Y$)**: For each available marker investigated as a mCoP, try to include it one by one as a covariate effect on the model parameters in $\mathcal{M}_0$. Collect all the BICc related to each model. Select the model (and related S) associated with the lowest BICc ($\mathcal{M}_*$).

3. **Conditional independence of the prospective mCoP and the intervention ($Y \perp A|S$)**: in $\mathcal{M}_*$, check that when adjusting jointly for the intervention effect and the mCoP, the intervention

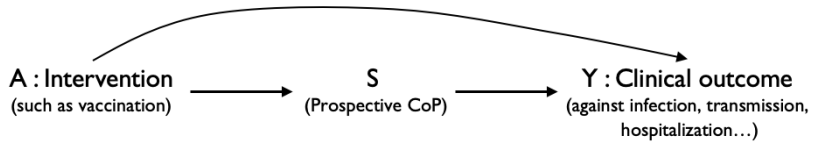


Figure 3.11 – Directed Acyclic graphs representing the definition of S as a CoP. According to Prentice's definition, S is a CoP if 1/ S must be correlated with the true outcome Y 2/ S must be affected by the intervention A 3/ The intervention A and outcome Y should be conditionally independent given the CoP: $Y \perp A|S$

effects turn out to be not statistically significant and the mCoP absorbs all the effect.

4. **Association between the intervention and the prospective mCoP ($A \rightarrow S$)**: in $\mathcal{M}_*$, check that the random effect associated with the parameter on which the mCoP is applied is decreased when adjusting for the mCoP in $\mathcal{M}_*$ rather than when adjusting for the intervention only in $\mathcal{M}_G$. This demonstrates that the mCoP captures more variability than the intervention does.

In our previous work, we modeled how the neutralization function of antibodies can prevent new infections and reduce viral dynamics. However, we overlooked a crucial aspect of mechanistic models: the role of viral load in stimulating new antibody production and enhancing the neutralizing response. Our simulations revealed that this oversight could lead to biases in estimating the vaccination's effect (CI-21-1). As a natural extension of this research, it is feasible to attempt to **establish a threshold for the mCoP against infection**. In pursuit of this goal, we leverage the reproductive number, as introduced in Section 3.2.2. This criterion serves as an indicator of whether the infection can sustain itself or not. Therefore, determining a threshold is possible by identifying the level of mCoP required to reduce the reproductive number below one with a high probability. This concept has previously been explored in collaboration with Marie Alexandre (HCI*-23-1). We are currently continuing to develop **joint models of virus and immune dynamics** and trying to bridge with mechanistic models across animal platforms, as detailed in the referenced future work Section 3.5. Finally, mCoP may be evaluated for other outcomes such as hospitalization or death [143], but this will necessitate human data instead of NHP data and likely an extension to joint models that include a survival-type outcome [47].

3.5 Perspectives

Below is a collection of perspectives, both short-term and long-term, that we aim to address by developing new within-host models. Ideas marked with an asterisk (*) are considered major in terms of challenges and potential impact.

Accounting for uncertainty and misspecification of models : As demonstrated in the previous chapter and referenced as RI-21-2, spline models are inferior to mechanistic models, particularly when aiming to predict system behavior beyond the learning period. A prospective avenue for further research on the analysis of ATI trials involves the adoption of mechanistic models instead of spline-based mixed models. In a starting work with my postdoctoral fellow Marie Alexandre, various mechanistic models of HIV rebound have been explored, including those presented in equations 3.1 and 3.3. However, none of these models perfectly reflect reality, and their selection may influence conclusions depending on the specific research question. Opting for model selection alone may overlook uncertainty in model choice, potentially leading to erroneous predictions and conclusions [22, 18]. Model averaging, which we privilege, entails integrating predictions from multiple models to enhance accuracy and reliability. Although literature on model averaging dedicated to mechanistic models in virus dynamics is scarce, except [76], Bayesian Model Averaging provides a Bayesian framework where model weights are determined by posterior model probabilities. Several approximation methods based on information criteria such as AIC or BIC exist for calculating posterior model probabilities [21]. However, consensus on the best criteria remains elusive, and other criteria, possibly based on the predictive abilities of the model, warrant exploration. Furthermore, exploring the use of Model Averaging for comparing the nAUC calculated with mechanistic models in ATI trials, as referenced in RI-21-2, holds promise. In summary, a key challenge will be to more systematically incorporate model averaging into analyses using mechanistic models.

Digital Twin and in sillico trials (*): A digital twin in the context of my work refers to a virtual representation or computational model that mirrors the immune system, viral infection processes, or the spread of infectious diseases within an individual. This sophisticated simulation tool integrates data from various sources to simulate and predict how the immune system interacts with pathogens under possible interventions. The primary goal of a digital twin in this domain is to provide a detailed, personalized model that can forecast and assess the efficacy of treatments or vaccines, and predict outcomes under various scenarios. In this field, very few article, except [113] describing a high-level roadmap, has been published. A first application that rely heavily on my existing work is the **application to optimization of ART in PLHIV**. Now that the on-off trials / short-cycle therapies have concluded and the results have been published ([44, 123], and [112] for a review), it is an opportune moment to author a retrospective paper. This paper will explore what could have been anticipated through modeling in term of effective strategies and what constitutes novel findings. The initiation of this work is planned under the supervision of an M1 or M2 student. If successful, the in silico trial pipeline that has been developed, capable of predicting a patient's specific response to ART therapy, can be regarded as a valuable digital twin for the quantitative guidance of ART optimization. A second application would focus on the **application to IL-7 immunotherapy optimization**.

To do so, an intriguing avenue to explore would involve employing reinforcement learning, comparing it with previously published methods such as optimal control [148] and Bayesian optimization (RI-18-2). My collaborators at Inria Flowers and I have already pursued a similar task in a different context, aiming to optimize the allocation of lockdown strategies during the COVID-19 pandemic (refer to Section 4.3.2). Given that the software, *Epidemioptim*, is already operational in Python, this could serve as a prospective internship project for a student at either the M1 or M2 level. Finally, a third example could be the **application to optimization of vaccine delivery**. We demonstrated that sex, age, geographic origins are specific determinants of vaccine efficacy. Investigation of circadian clock is undergoing [92]. Moreover, replication may be closely linked to adverse effects, particularly in immunosuppressed populations such as PLHIV. Finally, in an epidemic context with **ring vaccination** [82], the initial speed of establishment of the humoral response is a key factor for success. This raises a genuine question regarding which vaccine should be administered in which situation to a new individual with its own characteristics that we will try to address. If these three projects related to digital twins are conducted, it will enable the validation of models for virus dynamics, T cell proliferation, and B cell humoral response. Subsequently, an extension will be required on the modeling side to integrate these three aspects into a unified model. This integrated model will serve as the foundation for the pipeline designed for the creation of new studies or the in silico evaluation of intervention candidates. Thus, it will accelerate their development using a vaccinal digital twin.

Bridging between animals platforms and humans (*): Overall, the COVID-19 pandemic has highlighted the urgent need for tools to rapidly evaluate vaccine efficacy. Recently we started considering applying our methodologies to **rodent** platforms in collaboration with Veronique Godot at the Vaccine Research Institute. Mice experimentation offers a rapid and cost-effective avenue for evaluating the safety and efficacy of vaccine candidates prior to NHP and human trials. With their shorter lifespan, high reproductive rate, and genetic similarity to humans in key immune system functions, mice serve as invaluable models for studying vaccine-induced immune responses and predicting their translation to human populations. We have obtained data from an initial experiment, and we are currently investigating questions such as the optimal duration of follow-up needed to estimate parameters of humoral response in mice, such as the half-life of long-lived cells. Furthermore, the use of larger cohorts and a higher number of sacrifices in mice studies likely enables the observation of additional information regarding hidden compartments of mechanistic models. These trials may pave the way for observing populations of cells that do not circulate in the blood compartment, such as M cells in germinal centers. These investigations not only directly apply to our models but also open up potential for a broader research avenue to inform the design of mice experiments and suggest strategies for cross-species bridging from mice to NHP. Alongside a digital twin in mice and methods that may rely on allometric reasoning or other relevant mechanisms, it will become possible to more

effectively design NHP experiments. For instance, in order to ensure a productive and consistent infection pattern, animals are systematically infected by extremely large doses of virus, typically 10^6 PFU, while less than 10 infectious particles are sufficient to generate a natural infection. This implies for instance that NHP can be used to take "go/no go decisions" and discard drug candidates but may not be suited to more detailed analysis, such as the comparison of two drugs or establishment of mCoP threshold. Moreover, NHP cohorts exhibit high homogeneity in their levels of prospective mCoP due to identical vaccination and challenge dates. In summary, there is a recognized need to propose alternative designs that are statistically sound in NHP. Then, again, alongside a digital twin in NHP and methods that may rely on allometric reasoning or other relevant mechanisms, it will become possible to more effectively design human studies. In particular, a focus in the following years could be on Controlled Human Infection Model Study (CHIMS) [104], where volunteers are intentionally infected with pathogens under controlled conditions. Finally, unraveling the association between within-host parameters (in animals or humans) and transmission is only possible in experimental infection models [23]. In fact, basic questions, such as the role of viral load on transmission remain unknown for most viral infections. Similarly, understanding levels in the population that need to be reached for a mCoP to prevent infection or other outcome such as hospitalization or death is not an easy study to design (multi-scale modeling will be addressed in next section 4.4). An important challenge is to investigate if this type of study would be possible to design in rodents. In summary, we posit that a meticulous analysis of data from rodents (possibly mice), followed by NHP, and then humans, using dedicated mechanistic models built step by step on each platform, will accelerate the development of interventions. This pipeline from animal studies to clinical trials is standard, see [126] for the example of falvipiravir in Ebola. The ultimate aim is to achieve the capability to predict human outcomes (or at least part of it) solely based on mice studies.

Chapter 4

Outreach to Implementation : Evaluation in Population

4.1 Bibliometry

Finally, the third axis shifts the focus to the downstream aspects of development in population. It addresses the critical stages of population testing and the evaluation of intervention or vaccination strategies at a community or population level. This includes assessing the effectiveness of vaccination programs, understanding the dynamics of vaccine uptake, and evaluating the impact of various vaccination or interventions strategies. This comprehensive approach is essential for informing policy decisions and guiding effective public health interventions in the realm of vaccine development. Table 4.1 showcases the scientific outputs in this field. **This theme gave birth to 15 articles in peer-reviewed international journals, 12 international conferences, 3 national conferences, 6 seminars and 3 softwares or code releases. It is also funded by 2 national grants.** The interrelation of all these works is explained in the next subsections.

4.2 Cluster Randomized Trial for Evaluation of Interventions on Epidemics

This final research chapter concentrates on examining strategies (both vaccination and non-pharmaceutical interventions) employed in managing epidemics within populations. Assessing the impact of these interventions is complex due to the intricate nature of infectious disease progression, varying human behaviors and societal structures. A first method to evaluate the effect of an intervention in the population is **Cluster randomized trials (CRT)**, also known as a group-randomized trials.

Peer-reviewed International Journals		**First**	**Second**	**Before Last or Last**	**Other**
Lancet HIV					RI-17-4
Anaesthesia Critical Care & Pain medicine		RI-22-2			
American Journal of Public Health				RI-17-1; RI-17-2	
MDPI Covid/Viruses				RI-21-4	
Epidemics				RI-24-1	
Journal of Machine Learning Research				RI-21-1	
BMC Infectious Diseases				RI-23-3	
Journal of Clinical Microbiology					RI-16-5
R Journal		RI-17-5			
Statistical Methods in Medical Research				RI-19-2	
Biometrics		RI-16-2			
International Journal of Biostatistics				RI-23-1	
AIDS research			RI-18-1		
Observational Studies			RI-23-2		
Total (15)		3	2	8	2
International conferences	**Invited**	**Speaker**	**Mentee**	**Poster**	**Other**
CROI				HCI-23-5	
Society Mathematical Biology	HCI*-21-5				
PAGE		HCI-21-3			
Channel Network Conference	HCI*-21-1				
Society of clinical Trials	HCI*-17-1; HCI*-16-1	CI-15-1;			CI-19-1
Joint Statistical Meeting	HCI*-14-1				
Epidemics				CI-21-3	
Other		CI-16-1		CI-15-2	CI-18-3
Total (12)	5	3	0	2	2
National conferences	**Invited**	**Speaker**	**Mentee**	**Poster**	**Other**
ANRS MIE AC			CN-22-1		
JdS SFdS		CN-19-1; CN-16-1			
Total (3)	0	2	1	0	0
Seminars (6)	S-21-1; S-20-2; S-20-1; S-17-1; S-15-2; S-15-1				
International Grants (0)					
National Grants (2)	NG-21-1; NG-20-1				
Sofwares (3)	Soft-21-1; Soft-20-2; Soft-16-1				

Table 4.1 – Bibliometry for axis "Outreach to implementation : Evaluation in population"

They are particularly interesting because of prevention of **contaminations or interferences** [90]. In individual randomized trials, there are risks that individuals in the control group might indirectly benefit from the intervention given to the treatment group. For example, if a vaccine reduces the circulation of a pathogen in a community (herd immunity), this effect might extend to unvaccinated individuals, confounding the results. By randomizing at the cluster level, this issue can be mitigated.

Mixed-effects models [66] and **Generalized Estimating Equations** (GEE) [199] are statistical techniques used for analyzing correlated data, which is the case in CRT. In my work, we focused on the later because they require fewer distributional assumptions, such as distributions and independence between observations. GEE define the expectation of the dependent variables as a function of the independent variables and assume that the variance is a function of the mean. Additionally, they specify a working correlation structure separately for observations made on members of the same group. Finally, GEE provide a marginal intervention effect, whose population-averaged interpretation is preferred for making public health and policy decisions rather than the conditional, cluster-specific intervention effect estimated using mixed-effects models [157]. Of note, Marginal and conditional intervention effects are equal for identity and log links [167]. Together with a Elisabeth Turner, a colleague from Duke university, we wrote a review on good practices for CRT design (RI-17-1) and analysis (RI-17-2; HCI*-17-1).

In CRT analyzed with GEE (which is not a likelihood-based approach), missing data can significantly impact the validity of findings. In 2014, a review suggested that more than 90% of CRT do not handle properly missing data [48]. When data are **Missing at Random (MAR)** adequate statistical methods are mandatory to obtain unbiased results. Of note, MAR refers to a situation in data where the missingness is related to the observed data but not the unobserved data. In other words, the probability of a data point being missing is related to known variables, but not to the value of the missing data itself. **Weighting** and **multiple imputation** are two common methods for handling missing data. Weighting approach adjusts for missing data by giving more weight (using propensity score) to observed data that is similar to the missing data. Multiple imputation involves creating several complete datasets by replacing missing values with plausible estimates. These datasets are then analyzed separately, and the results are combined using Rubin's rule. Together with the same collaborator Elisabeth Turner, we compared the two approaches for CRT analysis (RI-16-2; CI-18-3; S-15-2). First, we demonstrated that clustering should not be accounted for when estimating the weights in this setting, even when there is clustering in the missingness mechanism. This counter-intuitive results is a mistake often made in practice and corroborated by [182]. Our simulations also showed that the performance of weighted GEE is comparable to and often faster than the Multiple imputation GEE approach. It is however to be nuanced since weighting approach are often very difficult to handle when missingness occurs jointly on multiple variables. See Section 4.4 for details

on how statistical methods for missing data adjustment in health data will continue to be developed within my research, in particular in the scope of predictions for ML.

Additional to missing data, CRT suffer from problems of **imbalance of baseline covariates**. It refers to the unequal distribution of participant characteristics across different clusters. This can occur due to the randomization at the cluster level rather than the individual level, leading to variations in important baseline characteristics between groups. GEE can incorporate covariates into the model in an outcome regression to control for imbalances. However, the results are subject to bias when the outcome model does not correspond to the true data generation process. On the flip side, the main limitation of weighting methods in GEE for CRT is the potential for bias if the propensity score for missingness are not accurately modeled and/or estimated. Thus we proposed a doubly robust GEE method for estimation in CRT (RI-17-5; CI-15-1; HCI*-16-1; HCI*-14-1; S-15-1). It was implemented in an R package (RI-19-2; ; CN-16-1; Soft-16-1). See Section 6.7 for the full version of the article presenting this work. In this article, we consider a study design in which a vector of P baseline covariates $X_{ij} = (X_{ij}^1, \ldots, X_{ij}^P)$ and outcome Y_{ij} are recorded for each subject $j = 1, \ldots, n_i$ in community $i = 1, \ldots, M$. Each outcome can be either observed $R_{ij} = 1$ or missing $R_{ij} = 0$. We want to make inference about the effect of a binary intervention A for which the probability of treatment attribution is denoted by p on the outcome Y by estimating the parameters $\beta = (\beta_0, \beta_A)^T$ indexing the marginal model $\mu_{ij}(\beta, A_i) = E(Y_{ij}|A_i) = \beta_0 + \beta_A A_i$, where $\mu_i(\beta, A_i) = [\mu_{ij}(\beta, A_i)]_{j=1,\ldots,n_i}$. The Doubly robust GEE writes as follow:

$$\sum_{i=1}^{M} \left[D_i^T V_i^{-1} W_i (Y_i - B_i) + \sum_{a=0,1} p^a (1-p)^{1-a} D_i^T V_i^{-1} (B_i - \mu_i) \right] = 0, \tag{4.1}$$

in which $D_i = \partial \mu_i / \partial \beta^T$ is the design matrix, V_i is the covariance matrix equal to $U_i^{1/2} C(\alpha) U_i^{1/2}$ with U_i a diagonal matrix with elements $var(y_{ij})$, and $C(\alpha)$ is the working correlation structure. This method is doubly robust because it remains consistent if either the propensity model ($W_i = diag(R_{ij}/P(R_{ij} = 1|X_i, A_i))$) or the outcome model ($B_i = E(Y_{ij}|A_i, X_i)$) is correctly specified, but not necessarily both. By correctly specified, we refer to a model that accurately corresponds to the true data generation process. The estimator provides a safeguard against model misspecification, enhancing the reliability of results in complex data scenarios. This was specifically applied to the TasP CRT which tested Antiretroviral Treatment as Prevention in Hlabisa sub-district, KwaZulu-Natal, South Africa (RI-17-4). Finally, I also got involved in analysis of trials in low-incomes setting in Africa: one comparing two measurement techniques for HIV infection in order to improve acceptance of testing (RI-16-5) and the other looking at using causal inference analysis to investigate the importance of rapid testing after infection (RI-18-1).

The spread of epidemics is significantly influenced by **network and connectivity patterns** within

a population. Networks, representing the social, physical, or geographical interactions among individuals, play a crucial role in determining how quickly and widely a disease can spread. High connectivity or densely connected networks facilitate rapid transmission of infectious diseases. The network features are quite never used as randomization covariates in CRT. Thus, **network features represent a source of imbalance in covariates**, see Figure 4.1. Together with JP onnela and his PhD student Patrick Staples from Havard School of Public health, with which I developed this project during my postdoctoral fellowship, we investigated which network features should be collected in priority to inform heterogeneity of the epidemics spread (RI-23-2; CI-16-1; CI-15-2; CN-19-1). We found that information on infection status of connected individuals or in the same network component are the one yielding to the best improvement in efficacy. Of note, no **time varying network features** is taken into account in this work. This could be a valuable extension of this work, but it is not straightforward. Although primarily a theoretical experiment, this project highlighted the significance of the network underlying an epidemic. This aligns with the findings of [177] and [77] regarding super-spreaders in the context of the COVID-19 pandemic. It also paves the way for innovative ideas in the design of **ring trials / reactive vaccination** particularly used in Ebola vaccination [82]. Refer to section 4.4 for further discussion on this topic.

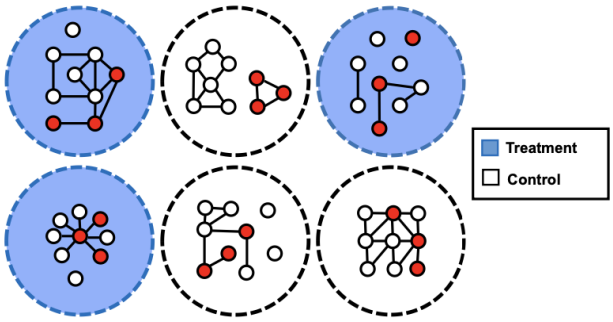


Figure 4.1 – Example of six clusters immediately after randomization in an idealized CRT. Three clusters have been randomized to treatment (blue), and three have been randomized to control (white). Each cluster has nine nodes, three of which are already infected at baseline and are shaded red. Internal network structure varies between clusters, and differing individual- and cluster-level covariates can be calculated for each cluster. For example, in the fifth (bottom-center) cluster, the mean degree is 2, the largest connected component size is 7. Individual outcomes in separate clusters are independent, as no edges exist between clusters (Figure from RI-23-2).

4.3 Modeling Between-host Dynamics of an Epidemics

4.3.1 Predicting the Effects of Non-Pharmaceutical Interventions

During the pandemics, I focused on **real-time observational data** provided within the **COVID-19 epidemics**. It refers to the data collected and analyzed continuously and promptly as the epidemic unfolds. The data used includes case reports, hospitalization data, testing data, vaccination data, climatic data and others. **Non-Pharmaceutical Interventions (NPI)** were crucial in controlling COVID-19. They included measures like social distancing, mask-wearing, hand hygiene, travel restrictions, and lockdowns. These interventions aimed to reduce virus transmission by limiting contacts between individuals. Predicting their effects involves understanding various epidemiological, social, and behavioral factors.

The core of **predicting NPI effects** lies in epidemiological models like the **SIR (Susceptible, Infected, Recovered)** model and its derivatives. These models, also called between hosts, simulate how the virus spreads through populations and how interventions might alter this spread. In our work, we focus on the SEIRAH model, see Equation 4.2 and Figure 4.2. The population of size N is divided into 5 compartments: susceptible S, latently exposed E, symptomatically infectious I, asymptomatic/pauci-symptomatically infectious A, hospitalized H, removed R (i.e., both recovered and deceased). Parameters D_X represent the average time in days spent in each compartment X. r_E and r_I respectively represent the proportion of symptomatic and hospitalized individuals. Finally, $\underline{b}$ is the most important parameter of transmission that is assumed reduced by α in asymptomatic individual due to reduced viral shedding.

$$\begin{cases} \frac{dS}{dt} = -\underline{b}\left(1-\frac{V}{N}\right)\frac{S(I+\alpha A)}{N} \\ \frac{dE}{dt} = \underline{b}\left(1-\frac{V}{N}\right)\frac{S(I+\alpha A)}{N} - \frac{E}{D_E} \\ \frac{dI}{dt} = \frac{r_E}{D_E}E - \frac{1-r_I}{D_Q}I - \frac{r_I}{D_I}I, \\ \frac{dR}{dt} = \frac{r_I I + A}{D_I} + \frac{H}{D_H} \\ \frac{dA}{dt} = \frac{1-r_E}{D_E}E - \frac{A}{D_I} \\ \frac{dH}{dt} = \frac{1-r_I}{D_Q}I - \frac{H}{D_H} \end{cases} \tag{4.2}$$

The particularity of our approach is not only considering one epidemics but multiple observation of the same epidemics process in multiple geo-localisation (region or departments), in a population framework. Noticing that these **epidemics models are mechanistic models**, we used the methods

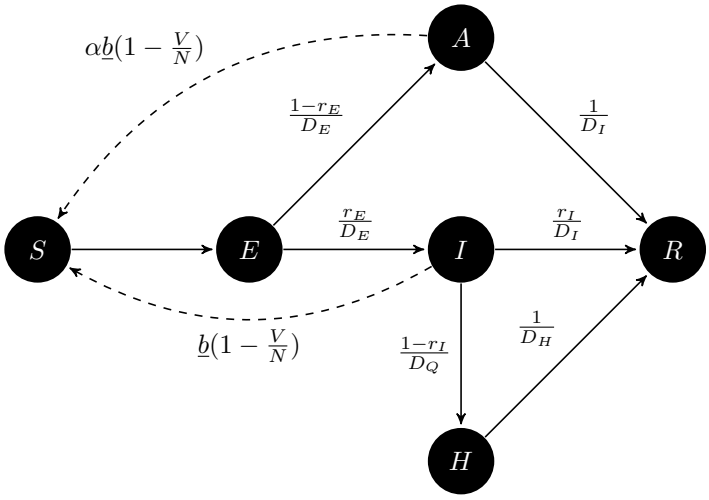


Figure 4.2 – SEIRAH model representation (Figure from RI-23-1)

mentioned in Section 2.2 and, in collaboration with Annabelle Collin from Inria team Monc, we fitted population Kalman filter estimation approaches (RI-23-1; HCI*-21-5; HCI-21-3;HCI*-21-1; S-20-1). See Section 6.8 for the full version of the article presenting this work. In this method of data assimilation, the parameters of the mechanistic model are treated as having dynamic behavior over time, which is modeled non-parametrically. Consequently, it becomes feasible to track the trajectory of transmission rate $\underline{b}$ across time and subsequently correlate it with the NPI being implemented using a regression method. Our findings indicate that all NPI considered in the study significantly contributed to the reduction of transmission rates. For example, the first lockdown reduced transmission by about 78% $[74\%;\ 82\%]$ and school closure reduced it by about 7% $[5\%;\ 8\%]$. Additionally, the study underscored the pronounced influence of weather conditions on disease transmission: rates decreased during summer months and escalated in winter. Moreover, an augmentation in disease transmissibility was observed with the emergence of VoC. A key goal of NPI is to reduce the effective reproduction number (Rt) (and not only transmission), which represents the average number of people to whom a single infected person will transmit the virus in a population where not everyone is susceptible. This concept has been introduced for within-host modeling in Section 3.2.2. Collaborating with Simon Cauchemez from Pasteur Institute and Juliette Paireau from Santé Publique France, we also proposed a two-step approach in which Rt is first estimated and NPI are regressed over the Rt (RI-23-3). Rt are derived using EpiEstim, which calculates the expected number of secondary infections [194]. Similar results were found as with the Kalman filters approach, however there was a disclaimer that confidence intervals found may have been too narrow. Finally, the study highlighted the importance of retrospective evaluation of interventions to **inform future decision-making**.

From a methodological perspective, the two studies depicted in the previous paragraph adopted

a **two-step approach** ($\underline{b}$ or Rt is estimated then regressed over NPI). Thus, they suffer from the weaknesses acknowledged in Section 2.2, i.e. the lack of propagation of uncertainty between the two steps as well as possible bias. To address this, we employed a **one-step strategy**, applying the SAEM estimation approach to the same dataset with a model of similar complexity (RI-24-1; HCI-23-5; CN-22-1). Again, this adjustment yielded results that were consistent with those obtained through the two-step process. Moreover, in this second study, we performed simulations that quantify benefits in term of public health. Without vaccines, the model predicted $159,000$ additional deaths and 1.48 million more hospitalizations in France. If a vaccine had been available within 100 days—a goal of the Coalition for Epidemic Preparedness Innovations (CEPI)—about one-third of deaths and three-quarters of hospitalizations could have been avoided. Finally, implementing a lockdown one week earlier could have prevented 20,000 deaths. In the search of always finding the most statistically accurate method to answer a relevant public health or medical question, we propose to compare one-step and two-step approaches as a first milestone for multi-scale modeling of epidemics (See Section 4.4).

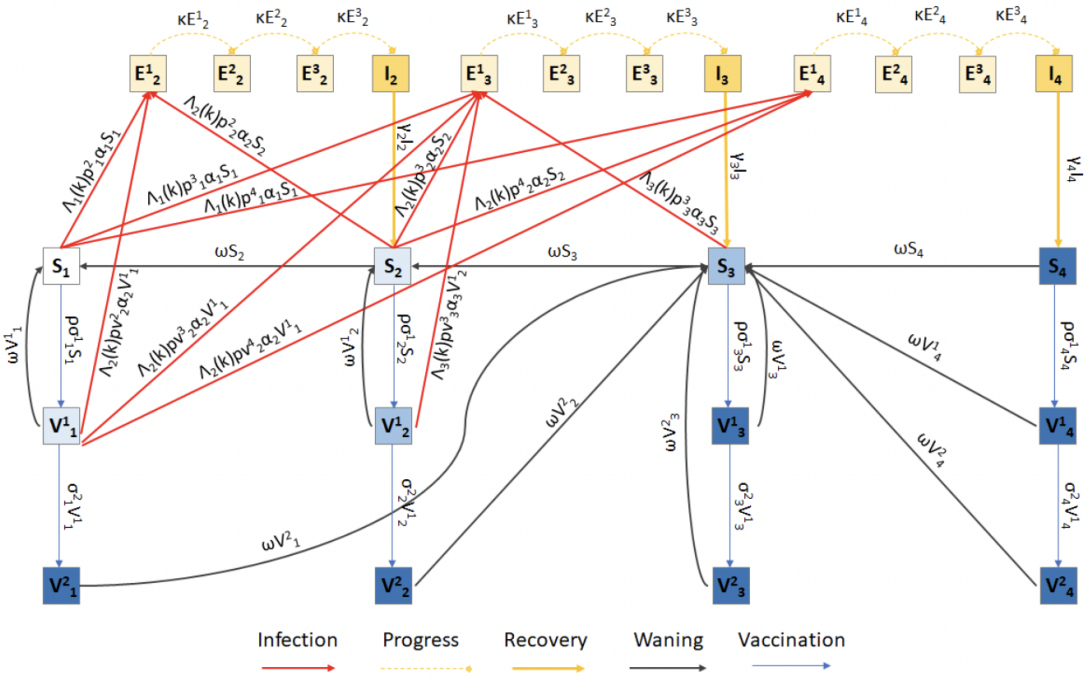


Figure 4.3 – Schematic of the age-structured SEIVS model for one age group. Parameters are not described in this manuscript and additional details can be found in RI-21-4 (Figure from RI-21-4).

As a last exemple of application, during the pandemic, there was a big challenge in forecasting, i.e. **short- mid- and long-term predictions** (weeks to several months) into the future, the effect of possible public health intervention. Collaborating with Jane Heffernan from York University, we

employed a more granular model briefly outlined in Figure 4.3 and expanded from [28]. This model is distinctive from the one presented before for two reasons: firstly, it includes **waning immunity** by following a SEIVS model structure (Susceptible-Exposed-Infectious-Vaccinated-Susceptible), and secondly, it incorporates **age stratification**. Such detailed models come at the cost of increased complexity and the need for fixing parameters or for more data to inform them. With this model, we projected the consequences of easing barrier gestures under varied assumptions about the longevity of immunity following infection and vaccination. The potential effects of a booster vaccine campaign were also examined (RI-21-4; CI-21-3). Of note, this work was published in mdpi journal after approval of publication policy by Inria. Nevertheless I will not reiterate this experience because I believe the reviewing standards were and still are very low quality. Regarding science, the findings from September 2020 emphasized the necessity of maintaining barrier gestures and enhancing vaccination efforts in particular in elderly. This result aligned with findings from other countries recommending the maintenance of NPI while vaccines were being rolled out [132]. These insights were important to informing the French situation, earning multiple citations from the French Scientific Council during the pandemic (S-20-2; Soft-20-2) as well as numerous press releases, see Section 1.3.13. To enhance the accuracy of estimating and predicting epidemic processes, it is crucial to deepen our understanding of the **connection between between-host dynamics and within-host dynamics**. This research perspective is further elaborated in Section 4.4.

4.3.2 Optimizing Lockdown Allocation

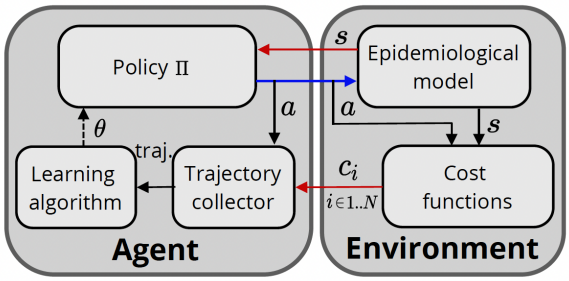


Figure 4.4 – Epidemic control as an optimization problem. s; a; c_i refer to environment states, control actions, and the i^{th} cost, while $traj.$ is their collection over an episode. Blue and red arrows match the input and output of the reinforcement learning function (Figure from RI-21-1).

The COVID-19 pandemic has underscored the significant impact of infectious diseases on public health and the economy, highlighting the need for **multi-objective cost optimization**. There have been multiple contribution concerning the optimization of intervention strategies for epidemic response. However, they mostly differ from their definitions of epidemiological models (SIR-type [212]; Agent-based models (ABM) [26]), of optimization methods (deterministic rules [189]; evo-

lutionary optimization [137]; bayesian optimization [26]) or of cost functions (for example [120] used length of school closure as well). In collaboration with Clément Moulin-Frier and researchers at Inria team Flowers, we decided to develop the EpidemiOptim toolbox which is a python platform allowing a simple definition of each of the aforementioned features. It facilitates interdisciplinary collaboration by integrating epidemiological models with **machine learning techniques** to optimize policy decisions. As an illustration, we got interested in finding the optimal allocation of lockdown measures during the COVID-19 pandemic trying to optimize both health costs and economic costs. We took the approach of **reinforcement learning** (RL) [186]: the control policy is seen as a learning agent, that interacts with an epidemiological model (SEIRAH) that is seen as its learning environment, see Figure 4.4. Each run of the epidemic is a learning episode, where the agent alternatively interacts with the epidemic through actions and observes the resulting states of the environment and their associated costs. Because the optimization is multi-objective, a Pareto front is built for decision (RI-21-1; S-21-1; Soft-21-1). See Section 4.4 for details on optimization of NPI or vaccination allocation that we envision.

4.4 Perspectives

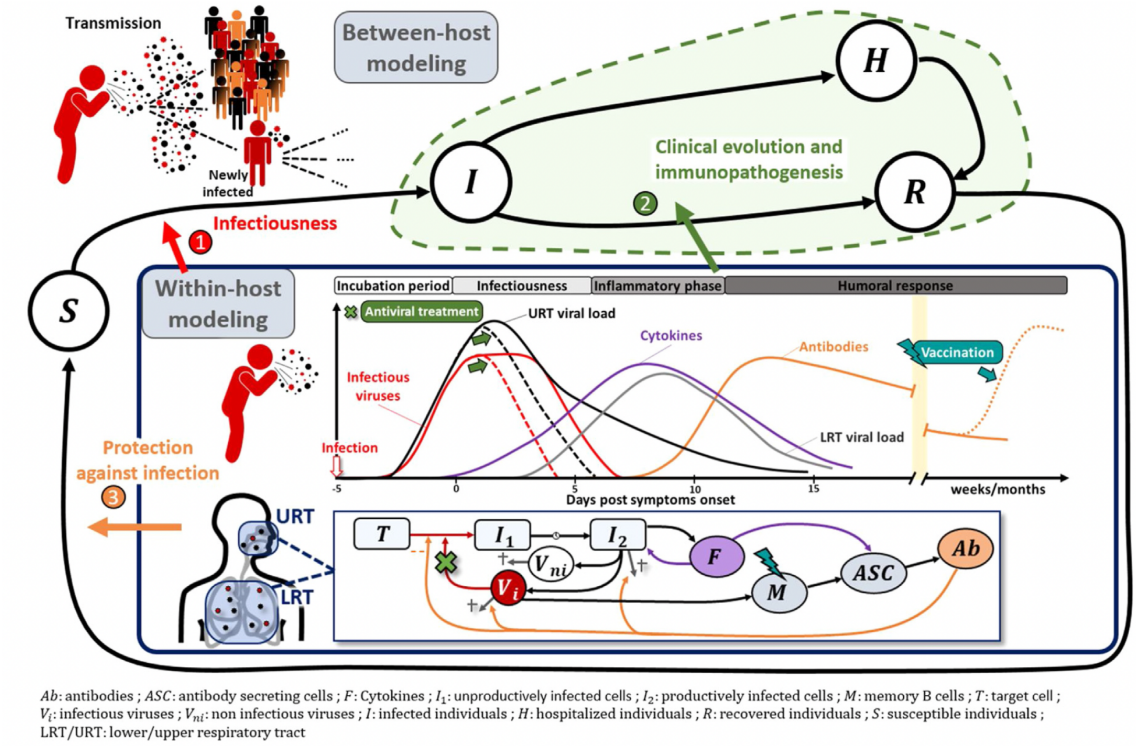
Below is a collection of perspectives, both short-term and long-term, that we think are major to advanced between-host modeling feeding it with within-host information. Ideas marked with an asterisk (*) are considered major in terms of challenges and potential impact.

Missing data in neural networks for predictions in Health data: Although not applied to CRT nor directly to mechanistic models, missing data presents a significant issue in all health data analysis. Particularly, in the project described in Section 2.6, which focuses on the use of ML / recurrent neural networks and specifically reservoir computing, addressing missing data is of high importance. In health data, it is rare for data points to be captured at precisely the same times, which complicates analysis. In collaboration with Xavier Hinault (Inria Mnemosyne), and Cécile Proust-Lima (Inserm Biostat), we plan to explore the application of classical Multiple Imputation (MI) methods for longitudinal data in our specific context, as reviewed by [91]. Specifically, we intend to concentrate on the Standard Fully Conditional Specification (FCS) [10] and Joint Multivariate Normal Imputation (JMNI)[173]. The FCS approach employs each time-varying covariate's repeated measurements as predictors in univariate imputation models, like chained equations. Conversely, JMNI employs data augmentation, a Markov chain Monte Carlo algorithm, assuming multivariate normality. This method, despite being potentially misspecified for binary and categorical data, has shown to perform adequately even for such variables, unless they exhibit severe skewness. Ultimately, we aim to compare these traditional approaches with innovative methods, particularly those based on

neural networks themselves. For example, Self-Attention-based Imputation for Time Series (SAITS) methods [56] have shown promise in managing missing time series data, especially when integrated with neural networks. Originating from natural language processing, SAITS employs self-attention mechanisms to identify temporal dependencies and patterns, offering a novel approach to handling missing data without dealing per-se with multiple imputation, which is yet to be investigated. Correct handling of missing information is a key challenge for future research in new ML methods.

Multi-scale Modeling of epidemics (*): In collaboration with Iris Ganser, a doctoral candidate at McGill University, we are evaluating one-step and two-step methods for estimating the effectiveness of NPI. Our planned analysis involves simulations under a variety of scenarios, incorporating ODE-based models and ABM for synthetic data creation. The objective is to demonstrate that the one-step approach may yield biased results, especially under conditions of rapid dynamic change and depletion of susceptibles in epidemics, and does not accurately convey uncertainty unless resampling techniques are employed (Sub-5). However, demonstrating that a one-step approach via mechanistic models outperforms others raises a concern: mechanistic models must be sufficiently complex to avoid misspecification, thereby necessitating a high number of parameters. Epidemiological studies will facilitate the calibration of some parameters, including changes in behaviors [69]. Introducing the evolution of pathogen virulence could also be of interest [4]. However, from a research perspective, our primary interest lies in integrating within-host modeling with between-host modeling. This multi-scale modeling provides a comprehensive picture of disease spread (RI-22-2), see Figure 4.5. Such models can predict how changes within individuals can affect the disease dynamics in a population, [183] for a theoretical example in multiple infections. This integration is particularly relevant for emerging infectious diseases and for understanding the impact of interventions such as drug development and vaccination, which is one of the focus of our group. To construct effective multi-scale models, our strategy encompasses two primary directions: Firstly, the development of new inference methods, as outlined in Section 2.6. These methods allowing estimation instead of calibration are likely to be based either on ODE models nested in ODE models [136, 166] or on ODE models nested within ABM [37, 190]. Moreover, incorporating more realistic contact networks into epidemiological models can improve our understanding of the emergence and spread of infectious diseases [103]. Secondly, the incorporation of innovative data to ascertain parameters not yet identified by existing research including CHIMS studies or very early cases reports [118]. This approach enables the examination of disease progression, immune responses, pathogen shedding, and the assessment of potential treatments or vaccines.

Reactive vaccination: This perspective concerns between-host optimization of vaccine delivery. Of note, it is different from within-host optimization of vaccine delivery presented in previous Section 3.5 for which RL could also be used. Actually, the EpidemiOptim toolbox presented in Section 4.3.2 of



Ab: antibodies ; *ASC*: antibody secreting cells ; *F*: Cytokines ; I_1: unproductively infected cells ; I_2: productively infected cells ; *M*: memory B cells ; *T*: target cell ; V_i: infectious viruses ; V_{ni}: non infectious viruses ; *I*: infected individuals ; *H*: hospitalized individuals ; *R*: recovered individuals ; *S*: susceptible individuals ; LRT/URT: lower/upper respiratory tract

Figure 4.5 – Models are used to understand the evolution of the viro-immunological response during an infection. As such, they can be used to disentangle the factors associated with (1) enhanced infectiousness during acute infection (2) progression towards a severe disease, and how to prevent it via pharmacological interventions (3) protection against infection via previous infection or vaccination (Figure from RI-22-2).

this chapter is now operational, presenting an exciting opportunity to extend its application to within-host problems, such as optimizing the allocation of IL-7 injections for PLHIV, as discussed in Section 3.2.4. Here, the intriguing avenue for exploration is optimizing vaccination policies and allocation through an age-structured epidemiological model, exemplified in Figure 4.3 or a network-structured spread exemplified in ring vaccination trials [82]. This type of optimization could be explored in terms of each dose's efficacy [131], as well as the vaccine's effectiveness regarding transmission, infection, or symptom reduction [187]. Furthermore, given the need to be reactive to adapt during emerging epidemics, it is imperative from a methodological standpoint to explore the application of Meta-RL. Meta-RL merges meta-learning with reinforcement learning principles, drawing inspiration from the human capacity to efficiently apply acquired knowledge to novel challenges [217]. This is achieved by designing the training process to include a variety of tasks/models, thereby encouraging the development of flexible learning strategies that can be quickly adapted or fine-tuned to new situations. By applying the algorithm across multiple epidemiological models informed by diverse epidemic parameters, Meta-RL can enable quicker adaptation to new emergent epidemics. In other

words, the approach allows an online adaptation in which dynamics model prior are combined with recent data and are rapidly adapted to the observed context.

Chapter 5

Conclusion

The preparation of this "Habilitation à Diriger des Recherches" has been a remarkable **period of reflection**, enabling me to **compile all my research work** from the start of my career in 2010. Although some might think it was completed "too late", as it occurred 8 years after I took my position as "Chargé de Recherche" at Inria, I believe it was timely. This delay allowed me to gain a **clear understanding of my accomplishments**, how they fit into the existing literature, and the research directions I wish to explore in the future. It has also allowed me to recognize weaknesses and begin to **develop improvement strategies**. Even though I am aware that research directions may vary depending on opportunities, I firmly believe that the perspectives mentioned in this manuscript will lead to thrilling scientific projects and collaborations.

In Chapters 2, 3 and 4, I have presented a complete overview of my research activities over the last 14 years. I made a deliberate choice to **separately discuss mathematical methods and biological applications** (within and between host), although in practice, this separation often does not exist. This decision was made to clearly outline both aspects: statistical analysis on one side, and modeling and interpretation of results in relation to the biological questions on the other. In reality, **these two facets are intertwined and complement each other as the research progresses**. My work fundamentally revolves around starting with a biological question that guides the selection of mathematical formalisms, leading to parallel mathematical/statistical and computational studies. These studies inform each other and contribute to answering the initial question, demonstrating an interplay between theory and application. I intend to continue enhancing my methodological development through the questions raised by the specific nature of the data collected and the question asked.

The pandemic has been a vibrant period for research in infectious diseases. It **fostered synergy in methodology**, driven by researchers' willingness to work with COVID-19 related data. However, it was a frustrating time, as the **pace of research was very different from that of policy-making**.

Nonetheless, this period led to multifaceted learning, which I believe has broadened my research to include between-host modeling - essential for adopting a comprehensive view of infectious disease modeling. The crisis also emphasizes the importance of collaboration among research groups in France that focus on infectious disease modeling and highlight the need for **effective communication and teamwork**. Our involvement has been formalized through active participation in the "Action Coordonnée Modélisation" initiative by ANRS MIE. I intend to keep these collaborations ongoing and strengthen them with additional research projects.

From an **international perspective**, we have also recently enhanced our **visibility** and participated in stimulating the community of modelers of within host in infectious diseases. For example, we will organize the next international workshop of this community in Bordeaux. The pandemic has also improved the team's capacity to collaborate worldwide - the "**Zoom era**" opens up prospects for easy remote collaboration. I intend to make the most of this in the coming years by **continuing to entertain my existing collaborations and build new ones**. Moreover, I believe that in recent years, the effort in our community to maintain **public data repositories** and **share code** via open access platforms such as GitHub has increased. I commit to applying these same principles of **open code and open resources** whenever possible in my research because I believe they are essential.

Biostatistics is pivotal in the domain of infectious diseases, acting as a fundamental element for comprehending the spread of diseases, the efficacy of interventions (both within and between host), and the shaping of public health policies. My personal fascination with these aspects explains the **broad range of applications** and research I pursue. However, I hope it is now clear that my specialization lies in the development of quantitative methods using mechanistic models, which serves as the primary connection among all my work. Looking ahead, I aim to maintain this broad spectrum and **enhance the impact of my research further**. In this regard, I believe it is crucial to **advance mechanistic models to the next level, enabling them to assimilate data from multiple sources and in large dimensions**. I hope this manuscript has provided clear research directions toward this goal.

Finally, I am very thankful for the research environment provided by the Inria/Inserm team SISTM and the Vaccine Research Institute, as well as the people within these groups. I hope we can continue to build strong connections, look in the same direction for applications, and **advance vaccine research for infectious diseases**.

Chapter 6

Selected Works (with Leader Contribution)

In this chapter, readers will find selected manuscripts from publications referenced in previous chapters. These articles have been chosen for their presentation of notable results or methodologies, or because they provide a detailed presentation of a series of results within a single manuscript. For each of these paper I was one of the major contributor either as first author or as senior mentor.

6.1 Clairon et al. 2023 (Computational Stat.) Optimal Control Estimation

Parameter estimation in nonlinear mixed effect models based on ordinary differential equations: an optimal control approach. Clairon Q., Pasin C. Balelli I., Thiébaut R. and Prague M.. *Computational Statistics.* in press - Oct 2023.

This article integrates in the first axis of my research "Statistical methodology for estimations and building of mechanistic models". It has been written in collaboration with a postdoctoral fellow I mentored.

I chose this paper as it exemplifies a key aspect of my research, which involves exploring available techniques for solving inverse problems and parameter estimation in patient-specific models, and then extending these to a population approach. This mirrors similar work I conducted on Kalman filters with A. Collin and P. Moireau. In this paper, our focus is on the optimal control approach. We describe the methodology and test it through simulations in various scenarios : 1. presence of model error (hypoelliptic stochastic model); 2. partially observed framework with unknown initial conditions (glucose and insulin regulation model); and 3. presence of poorly identifiable parameters (antibody concentration evolution). A significant challenge was deriving the estimators' variance.

Computational Statistics
https://doi.org/10.1007/s00180-023-01420-x

ORIGINAL PAPER

Parameter estimation in nonlinear mixed effect models based on ordinary differential equations: an optimal control approach

Quentin Clairon[1,2] · **Chloé Pasin**[3] · **Irene Balelli**[4] · **Rodolphe Thiébaut**[1,2] · **Mélanie Prague**[1,2]

Received: 10 December 2022 / Accepted: 18 September 2023

Abstract

We present a method for parameter estimation for nonlinear mixed-effects models based on ordinary differential equations (NLME-ODEs). It aims to regularize the estimation problem in the presence of model misspecification and practical identifiability issues, while avoiding the need to know or estimate initial conditions as nuisance parameters. To this end, we define our estimator as a minimizer of a cost function that incorporates a possible gap between the assumed population-level model and the specific individual dynamics. The computation of the cost function leads to formulate and solve optimal control problems at the subject level. Compared to the maximum likelihood method, we show through simulation examples that our method improves the estimation accuracy in possibly partially observed systems with unknown initial conditions or poorly identifiable parameters with or without model error. We conclude this work with a real-world application in which we model the antibody concentration after Ebola virus vaccination.

Keywords Dynamic population models · Ordinary differential equations · Optimal control theory · Mechanistic models · Nonlinear mixed effects models · Clinical trial analysis

✉ Quentin Clairon
quentin.clairon@u-bordeaux.fr

[1] Inria Bordeaux Sud-Ouest, Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR1219, University of Bordeaux, 33000 Bordeaux, France

[2] Vaccine Research Institute, 94000 Créteil, France

[3] Department of Infectious Diseases and Hospital Epidemiology, University Hospital, Collegium Helveticum, Institute of Medical Virology, University of Zurich, Zurich, Switzerland

[4] Centre Inria d'Université Côte d'Azur, EPIONE Research Project, Valbonne, France

Published online: 14 October 2023

1 Introduction

Ordinary differential equation (ODE) models are standard in population dynamics, epidemiology, virology, pharmacokinetics, or genetic regulatory network analysis since they can describe the main mechanisms of interaction between different biological components of complex systems and their evolution over time and because they also provide reasonable approximations to stochastic dynamics (Perelson et al. 1996; Engl et al. 2009; Villain et al. 2019).

For experimental designs with a large number of subjects and a limited number of individual measurements, nonlinear mixed-effects models may be more relevant than single-subject models, since they allow to collect information from the entire population while accounting for variability among individuals. For instance, clinical trials and pharmacokinetics/pharmacodynamics studies often fall into this category (Lavielle and Mentre 2011; Guedj et al. 2007). Formally, we are interested in a population where the dynamic of each subject $i = 1, \dots n$ is modeled by the d-dimensional ODE:

$$\begin{cases} \dot{x}_i(t) = f_{\theta,b_i}(t, x_i(t)) \\ x_i(0) = x_{i,0} \end{cases} \tag{1}$$

where f is a $d-$dimensional vector field, θ is a $p-$dimensional parameter, $b_i \sim N(0, \Psi)$ is a $q-$dimensional random effect where Ψ is a variance-covariance matrix, $x_{i,0}$ is the initial condition for subject i. We denote $X_{\theta,b_i,x_{i,0}}$ the solution of (1) for a given set $\left(\theta, b_i, x_{i,0}\right)$. In (1), we can also incorporate covariate functions z_i which are omitted here for the purpose of clarity.

Our goal is to estimate the true population parameters (θ^*, Ψ^*) as well as the true subject specific realizations $\left\{b_i^*\right\}_{i=1,\dots n}$ from partial and noisy observations coming from n subjects and described by the following observational model:

$$y_{ij} = CX_{\theta^*,b_i^*,x_{i,0}^*}(t_{ij}) + \epsilon_{ij}.$$

For the i-th subject, we denote t_{ij} its j-th measurement time-point on the observation interval $[0, T]$ and n_i its total number of available measurements. Here C is a $d^o \times d$ sized observation matrix emphasizing the potentially partially observed nature of the process and $\epsilon_{ij} \sim \sigma^* \times N(0, I_{d^o})$ is the measurement error. The vector $\mathbf{y_i} = \left\{y_{ij}\right\}_{j=1,\dots n_i}$ corresponds to the set of observations available for the i-th subject and $\mathbf{y} = \left\{\mathbf{y_i}\right\}_{i=1,\dots n}$ is the set of all observations in the population. We also assume that only a subset $x_{i,0}^{k*}$ of $x_{i,0}^*$ is perfectly known, the other ones, denoted $x_{i,0}^{u*}$, being unknown and they are ordered as follows $x_{i,0} = \left(\left(x_{i,0}^u\right)^T, \left(x_{i,0}^k\right)^T\right)^T$ for the sake of clarity. Nonetheless, pre-existing information can be available for $x_{i,0}^{u*}$ under the form of a priori distribution with a possibly parameter dependent density $\mathbb{P}\left(x_{i,0}^u \mid \theta, \Psi, b_i\right)$. The same holds for (θ, Ψ) for which a priori distribution $\mathbb{P}(\theta, \Psi)$ can be available.

Our problem belongs to the class of parameter estimation problem in nonlinear mixed effect models based on ODEs (NLME-ODEs). In this context, frequentist methods based on likelihood maximization (via different numerical procedures: Laplace approximation (Pinheiro and Bates 1994), Gaussian quadrature (Guedj et al. 2007) or SAEM (Comets et al. 2017; Lavielle and Mentré 2007)) and Bayesian ones aiming to reconstruct the a posteriori distribution or to derive the maximum a posteriori estimator (via MCMC algorithms (Lunn et al. 2000; Huang and Dagne 2011), importance sampling (Raftery and Bao 2010), approximation of the asymptotic posterior distribution (Prague et al. 2013)) have been proposed. In particular, dedicated methods/softwares using the structure of ODE models have been implemented to increase numerical stability and speed up convergence rate (Tornoe et al. 2004), to reduce the computational time (Donnet and Samson 2006) or to avoid the repeated model integration and estimation of initial conditions (Wang et al. 2014). However, all the preceding methods face similar pitfalls due to specific features of population models based on ODEs (with the exception of (Wang et al. 2014)):

1. They do not take into account the presence of model misspecification, a common feature of ODE models used in biology. Indeed, the ODE modeling process suffers from model inadequacy, understood as the discrepancy between the mean model response and the real process, as well as residual variability subject to specific stochastic perturbations or missing elements that disappear by averaging over the entire population (Kennedy and O'Hagan 2001). As examples of the causes of model inadequacies, one can think of the ODE models used in epidemiology and virology, which are derived by approximations in which, for example, interactions are modeled by pairwise products, while higher order terms and/or the influence of unknown/unmeasured external factors are neglected. As for residual variability, recall that biological processes are often stochastic and the justification for deterministic modeling lies in the approximation of stochastic processes (Kurtz 1978; Kampen 1992). Moreover, in the context of NLME-ODEs, new sources of model uncertainties emerge. Firstly, error measurement in covariates z_i can lead to use a proxy function $\widehat{z}_i$ instead of z_i (Huang and Dagne 2011). Secondly, the sequential nature of most inference methods leads to estimate $\left\{b_i^*\right\}_{i=1,\dots n}$ based on an approximation $\widehat{\theta}$ instead of θ^*. Thus, the structure of mixed-effect models spreads measurement uncertainty into the mechanistic model structure during the estimation. It turns classical statistical uncertainties into model error causes. Estimation of θ^*, Ψ^* and $\left\{b_i^*\right\}_{i=1,\dots n}$ must be performed in the presence of the model error, although it is known to dramatically affect the accuracy of methods that do not take it into account (Brynjarsdottir and O'Hagan 2014).
2. They have to estimate or make assumptions on the true unknown initial conditions $x_{i,0}^{u*}$. In ODE models, the initial conditions are generally nuisance parameters in the sense that inferring their values does not bring answers to the scientific questions which motivate the model construction but is necessary for the estimation of the relevant parameters. For example, partially observed compartmental models used in pharmacokinetics/pharmacodynamics often

involve unknown initial conditions which needs to be inferred to estimate the transmission rates between compartments, which are the true parameters of interest. Unknown initial conditions imply either assumptions on $x_{i,0}^{u*}$ values (Lavielle and Mentre 2011), another potential cause of model misspecifications, or their estimation (Huang and Lu 2008). This latter case requires a priori knowledge to derive $\mathbb{P}\left(x_{i,0}^{u} \mid \theta, \Psi, b_i\right)$ expression and simultaneous inference of $(b_i^*, x_{i,0}^{u*})$ as subject specific parameters. This increases the complexity of the related optimization problem and can degrade estimation accuracy.

3. They can face accuracy degradation when the inverse problem of parameter estimation is ill-posed (Engl et al. 2009) due to practical identifiability issues. Ill-posedness in ODE models is often due to the geometry induced by the mapping $\left(\theta, b_i, x_{i,0}\right) \longmapsto CX_{\theta, b_i, x_{i,0}}$, where there can be a small number of relevant directions of variation skewed from the original parameter axes (Gutenkunst et al. 2007). This problem, called sloppiness, often appears in ODE models used in biology (Gutenkunst et al. 2007; Leary et al. 2015) and leads to an ill-conditioned Fisher Information Matrix. For maximum likelihood estimators this is a cause of high variance due to the Cramér-Rao bound. For Bayesian inference, it leads to a nearly singular asymptotic posterior distribution because of Bernstein–von Mises theorem (see Campbell (2007) for the computational induced problems). Although this problem is in part mitigated in NLME-ODEs which merge different subjects for estimating (θ^*, Ψ^*) and use distribution of $b_i \mid \Psi$ as prior at the subject level (Lavielle and Aarons 2015), estimation accuracy can benefit from the use of regularization techniques.

These specific features of ODE-based population models limit the amount of information classic approaches can extract for estimation purposes from observations no matter their qualities or abundances. This advocates for the development of new estimation procedures. Approximate methods (Varah 1982; Ramsay et al. 2007) have already proven to be useful for ODE models to face such issues. They rely on an approximation of the solution of the original ODE (1) which is expected to have a smoother dependence with respect to the parameters and to relax the constraint imposed by the model during the estimation process. The interest of such approximations is twofold. Firstly, they produce estimators with a better conditioned variance matrix comparing to classic likelihood based approaches. Secondly, they reduce the effect of model error on estimator accuracy. Also, some of these approximations bypass the need to estimate initial conditions (Ramsay et al. 2007; Clairon 2020). Still, these methods are currently limited to cases where observations are coming from one subject.

In this work, we develop a new estimation method adapted to NLME-ODEs integrating such approximations to mitigate the effect of model misspecification and poorly identifiable parameters on estimation accuracy, while avoiding the need to estimate $x_{i,0}^{u*}$ as additional subject specific parameters. At the contrary to the mentioned methods, we propose here a hierarchical profiling approach, taking the form of a nested estimation procedure, instead of relying on a classic joint likelihood

specification. One point being to avoid high-dimensional integrations often required by likelihood approaches in mixed-effect models (Pinheiro and Bates 1994). The unknown initial conditions $\left\{x_{i,0}^{u*}\right\}_{i=1,\dots n}$ are seen as nuisance parameters for $\left\{b_i^*\right\}_{i=1,\dots n}$ estimation, which are in turn considered as nuisance parameters for population parameter $(\theta^*, \Psi^*, \sigma^*)$ estimation. This leads to the construction of outer, middle and inner criteria for the estimation of $(\theta^*, \Psi^*, \sigma^*)$, $\left\{b_i^*\right\}_{i=1,\dots n}$ and $\left\{x_{i,0}^{u*}\right\}_{i=1,\dots n}$ respectively. The inner criteria is designed to incorporate $\mathbb{P}\left(x_{i,0}^{u} \mid \theta, \Psi, b_i\right)$ if an expression is proposed for it but can also be defined without if no prior information exists for $\left\{x_{i,0}^{u*}\right\}_{i=1,\dots n}$. Also, this criterion accounts for model error presence by assuming that the actual dynamic of each subject is better described by a perturbed version of the ODE (1). This added perturbation aims to capture various sources of errors at the subject level (Brynjarsdottir and O'Hagan 2014; Tuo and Wu 2015). We control the magnitude of the acceptable perturbations by defining the inner criteria through a cost function balancing the two contrary objectives of fidelity to the observations and to the original model: to this end, we introduce a model discrepancy penalization term. The practical computation of the chosen perturbations requires to solve optimal control problems (Clarke 2013) known as tracking problems. This is done using a method inspired by Cimen and Banks (2004) which has the advantage to automatically provide an estimator for $x_{i,0}^{u*}$ with no additional computational costs. This is the key element to efficiently profile on unknown initial conditions during b_i^* estimation, and treat them as nuisance parameters instead of integrating them into b_i definition, as it is usually done in the previously mentioned methods.

In Sect. 2, we present the inner, middle and outer criteria used to define our estimator. In Sect. 3, we compare our approach with classic maximum likelihood in simulations. Then, we proceed to the real data analysis coming from clinical studies and a model of the antibody concentration dynamics following immunization with an Ebola vaccine in East African participants (Pasin et al. 2019). Section 5 concludes and discuss future extensions of the method.

2 Estimator construction: definition of the inner/middle/outer criteria

From now on, we use the Cholesky decomposition $\sigma^2\Psi^{-1} = \triangle^T\triangle$ and the parametrization $\phi := (\theta, \Delta, \sigma)$ instead of (θ, Ψ, σ) to enforce positiveness and symmetry of Ψ and denote in a summarized way the set of all population parameters. The norm $\|.\|_2$ denotes the classic Euclidean one defined by $\|b\|_2 = \sqrt{b^T b}$. We estimate the population and individual parameters via a nested procedure:

- Estimator $\widehat{\phi}$ obtained by minimization of an outer criterion F based on an approximation of $\min_b \min_{x_0^u} \left(-\ln \mathbb{P}\left(\phi, b, x_0^u \mid \mathbf{y}\right)\right)$, the log joint-distribution of $\left(\phi, b, x_0^u\right)$ sequentially profiled on $b := \left\{b_i\right\}_{i=1,\dots n}$ and $x_0^u := \left\{x_{i,0}^u\right\}_{i=1,\dots n}$, which are respectively the set of all random effects and unknown initial conditions among all subjects.
- Estimator $\widehat{b}_i := \widehat{b}_i(\phi)$ obtained for each subject i by minimization of a middle criterion G_i based on an approximation of $\min_{x_{i,0}^u} \left(-\ln \mathbb{P}(\mathbf{y_i}, b_i, x_{i,0}^u \mid \phi)\right)$, the log joint-distribution of the data, the random effects and unknown initial conditions profiled on the latter.
- Estimator $\widehat{x_{i,0}^u} := \widehat{x_{i,0}^u}(\phi, b_i)$ obtained for each subject i by minimization of an inner criterion H_i based on an approximation of $-\ln \mathbb{P}(\mathbf{y_i}, x_{i,0}^u \mid \phi, b_i)$, the log joint-distribution of the data and unknown initial conditions.

Our estimation procedure can be expressed in a pseudo-algorithmic way.

1. Outer criteria minimization:

$$\begin{aligned}\widehat{\phi} &= \arg\min_{\phi} F(\phi)\\ &:= \arg\min_{\phi} \min_{b} \min_{x_0^u} -2\ln \widetilde{\mathbb{P}}(\phi, b, x_0^u \mid \mathbf{y})\\ &= \arg\min_{\phi} -2\ln \widetilde{\mathbb{P}}(\phi, \widehat{b}, \widehat{x_0^u} \mid \mathbf{y}),\end{aligned}$$

for a given subject i and ϕ value:

2. Middle criteria minimization:

$$\begin{aligned}\widehat{b}_i(\phi) &= \arg\min_{b_i} G_i(b_i \mid \phi)\\ &:= \arg\min_{b_i} \min_{x_{i,0}^u} -2\ln \widetilde{\mathbb{P}}(\mathbf{y_i}, b_i, x_{i,0}^u \mid \phi)\\ &= \arg\min_{b_i} -2\ln \widetilde{\mathbb{P}}(\mathbf{y_i}, b_i, \widehat{x_{i,0}^u} \mid \phi),\end{aligned}$$

for a given b_i value:

3. Inner criteria minimization:

$$\begin{aligned}\widehat{x_{i,0}^u}(\phi, b_i) &= \arg\min_{x_{i,0}^u} H_i(x_{i,0} \mid \phi, b_i)\\ &:= \arg\min_{x_{i,0}^u} -2\ln \widetilde{\mathbb{P}}(\mathbf{y_i}, x_{i,0}^u \mid \phi, b_i).\end{aligned}$$

In the following sections, we derive the expressions of F, G_i and H_i starting with H_i since each criterion construction rely on lower level ones. Finally, despite that the following formal presentation of criteria are made for any $\mathbb{P}\left(x_{i,0}^u \mid \phi, b_i\right)$ expressions, we have to restrict ourselves to uniform, normal and log-normal densities in practice to use our numerical procedures.

2.1 Inner criterion

In this section, we construct the criteria H_i used to estimate $x_{i,0}^{u*}$ for a given (ϕ, b_i) value. A classic procedure would lead to jointly estimate $\left(b_i^*, x_{i,0}^{u*}\right)$ by maximization of the log joint-likelihood function of the data and $(b_i, x_{i,0}^u)$. However for each subject, we want to:

1. profile on $x_{i,0}^{u*}$ during random effects estimation to limit b_i^* estimation degradation due to presence of nuisance parameters,
2. use prior knowledge given by $\mathbb{P}\left(x_{i,0}^u \mid \phi, b_i\right)$ if available,
3. allow an acceptable deviation from the assumed model at the population level to take into account possible model misspecifications.

To solve the first and second point, we define our estimator:

1. as the maximizer of the joint conditional likelihood $\mathbb{P}(\mathbf{y_i}, x_{i,0}^u \mid \phi, b_i)$ if $\mathbb{P}\left(x_{i,0}^u \mid \phi, b_i\right)$ is available,
2. otherwise as the maximizer of $\mathbb{P}(\mathbf{y_i} \mid \phi, b_i, x_{i,0}^u)$.

From the expression $\mathbb{P}(\mathbf{y_i}, x_{i,0}^u \mid \phi, b_i) = \mathbb{P}(\mathbf{y_i} \mid \phi, b_i, x_{i,0}^u)\mathbb{P}\left(x_{i,0}^u \mid \phi, b_i\right)$, we derive $\arg\max_{x_{i,0}^u} \mathbb{P}(\mathbf{y_i} \mid \phi, b_i, x_{i,0}^u) = \arg\max_{x_{i,0}^u} \mathbb{P}(\mathbf{y_i}, x_{i,0}^u \mid \phi, b_i)$ if $\mathbb{P}\left(x_{i,0}^u \mid \phi, b_i\right)$ is constant. So, the estimation criteria in absence of prior information is equivalent to choosing a uniform prior over $x_{i,0}^u$ space and constitute only a particular case. We will thus focus on $\mathbb{P}(\mathbf{y_i}, x_{i,0}^u \mid \phi, b_i)$ from now one. We have:

$$\begin{aligned}\mathbb{P}(\mathbf{y_i}, x_{i,0}^u \mid \phi, b_i) &= \prod_j \mathbb{P}(y_{ij} \mid \phi, b_i, x_{i,0}^u)\mathbb{P}\left(x_{i,0}^u \mid \phi, b_i\right) \\ &= \prod_j (2\pi)^{-d^o/2}\sigma^{-d^o} e^{-0.5\left\|CX_{\theta,b_i,x_{i,0}}(t_{ij})-y_{ij}\right\|_2^2/\sigma^2}\mathbb{P}\left(x_{i,0}^u \mid \phi, b_i\right),\end{aligned}$$

from which we derive the joint likelihood estimator:

$$\begin{aligned}\widehat{x_{i,0}^u}(\phi, b_i) &= \arg\min_{x_{i,0}^u} -2\ln \mathbb{P}(\mathbf{y_i}, x_{i,0}^u \mid \phi, b_i) \\ &= \arg\min_{x_{i,0}^u} \left\{ \frac{1}{\sigma^2}\sum_j \left\|CX_{\theta,b_i,x_{i,0}}(t_{ij}) - y_{ij}\right\|_2^2 - 2\ln \mathbb{P}\left(x_{i,0}^u \mid \phi, b_i\right) \right\}.\end{aligned}$$

We also want to allow the presence of perturbations at the subject scale comparing to the original model defined at the population level. For this, we assume the regression function is no longer $X_{\theta,b_i,x_{i,0}}$, but rather $X_{\theta,b_i,x_{i,0},u_i}$, the solution of:

$$\begin{cases} \dot{x}_i(t) = f_{\theta,b_i}(t, x_i(t)) + Bu_i(t) \\ x_i(0) = x_{i,0}. \end{cases} \tag{2}$$

This perturbed ODE has been obtained by the addition of the forcing term $t \mapsto Bu_i(t)$ to ODE (1) with B a $d \times d_u$ matrix and u_i a function in $L^2\left([0,T], \mathbb{R}^{d_u}\right)$ representing the perturbation. However, to ensure the possible perturbations remain small, we replace the data fitting criterion $\sum_j \left\| CX_{\theta,b_i,x_{i,0}}(t_{ij}) - y_{ij} \right\|_2^2$ by $\min_{u_i} \mathcal{C}_i(x^u_{i,0}, u_i \mid \theta, b_i, U)$, where

$$\mathcal{C}_i(x^u_{i,0}, u_i \mid \theta, b_i, U) = \sum_j \left\| CX_{\theta,b_i,x_{i,0},u_i}(t_{ij}) - y_{ij} \right\|_2^2 + \left\| u_i \right\|^2_{U,L^2},$$

and $\left\| u_i \right\|^2_{U,L^2} = \int_0^T u_i(t)^T U u_i(t) dt$ is the weighted Euclidean norm. Here, the magnitude of the allowed perturbations is controlled by a positive definite and symmetric weighting matrix U. Finally, we obtain:

$$\widehat{x^u_{i,0}}(\phi, b_i) := \arg\min_{x^u_{i,0}} H_i(x^u_{i,0} \mid \phi, b_i) \tag{3}$$

where

$$H_i(x^u_{i,0} \mid \phi, b_i) = \min_{x^u_{i,0}} \left\{ \frac{1}{\sigma^2} \min_{u_i} \mathcal{C}_i(x^u_{i,0}, u_i \mid \theta, b_i, U) - 2 \ln \mathbb{P}\left(x^u_{i,0} \mid \phi, b_i\right) \right\}.$$

Computing $H_i(x^u_{i,0} \mid \phi, b_i)$ requires to solve the infinite dimensional optimization problem $\min_{u_i} \mathcal{C}_i(x^u_{i,0}, u_i \mid \theta, b_i, U)$ in $L^2\left([0,T], \mathbb{R}^{d_u}\right)$. This problem belongs to the field of optimal control theory for which dedicated approaches have been developed (Sontag 1998; Aliyu 2011; Clarke 2013). Here we use the same method as in Clairon (2020) which is detailed in Appendix A. All it requires from the user is to specify a pseudo-linear representation of ODE (1), i.e., a possibly state-dependent matrix $A_{\theta,b_i}\left(t, x_i(t)\right)$ and state-independent vector $r_{\theta,b_i}(t)$ such that:

$$f_{\theta,b_i}(t, x_i(t)) = A_{\theta,b_i}\left(t, x_i(t)\right) x_i(t) + r_{\theta,b_i}(t). \tag{4}$$

This formulation is crucial for solving the optimal control problem in a computationally efficient way. Linear models already fit in this formalism with $A_{\theta,b_i}(t) := A_{\theta,b_i}\left(t, x_i(t)\right)$. For nonlinear models, the pseudo-linear representation is not unique but always exists (in order to exploit this non-uniqueness as an additional degree of freedom, see Cimen (2008) section 6). This method presents the advantage of formulating $\min_{u_i} \mathcal{C}_i(x^u_{i,0}, u_i \mid \theta, b_i, U)$ as a quadratic form (or a sequence of quadratic forms) with respect to $x^u_{i,0}$. Thus, if we choose a uniform, normal or log-normal law for $\mathbb{P}\left(x^u_{i,0} \mid \phi, b_i\right)$, $\arg\min_{x^u_{i,0}} H_i(x^u_{i,0} \mid \phi, b_i)$ has a closed form expression (approximated for log-normal), and obtaining $\widehat{x^u_{i,0}}(\phi, b_i)$ does not add any computational complexity comparing to $\min_{u_i} \mathcal{C}_i(x^u_{i,0}, u_i \mid \theta, b_i, U)$.

For a given $x^u_{i,0}$, the perturbation u_i corresponding to the solution of $\min_{u_i} \mathcal{C}_i(x^u_{i,0}, u_i \mid \theta, b_i, U)$ is named optimal control and denoted $\overline{u}_{i,\phi,b_i,x^u_{i,0}}$. In particular,

we denote $\overline{u}_{i,\theta,b_i} := \overline{u}_{i,\theta,b_i,\widehat{x_{i,0}}}$ the optimal control corresponding to the initial condition estimator $\widehat{x_{i,0}} = \left(\widehat{x^u_{i,0}}(\phi, b_i)^T, \left(x^k_{i,0}\right)^T\right)^T$. The solution of (2) corresponding to the optimal control $u_i := \overline{u}_{i,\theta,b_i}$ is denoted $\overline{X}_{\theta,b_i}$ and named optimal trajectory: it will be considered as the regression function for the i-th subject. $\overline{X}_{\theta,b_i}$ is thus defined as solution of ODE (2) which needs the smallest perturbation in order to get close to the observations. In particular, $\overline{X}_{\theta,b_i}$ and $\overline{u}_{i,\theta,b_i}$ are respectively the subject specific state variable and perturbation such that:

$$H_i(\widehat{x^u_{i,0}}(\phi, b_i)|\phi, b_i) = \frac{1}{\sigma^2}\left\{\sum_j \left\|C\bar{X}_{\theta,b_i}(t_{ij}) - y_{ij}\right\|_2^2 + \left\|\bar{u}_{i,\theta,b_i}\right\|^2_{U,L^2}\right\} - 2\ln \mathbb{P}(\widehat{x^u_{i,0}}(\phi, b_i)|\phi, b_i). \tag{5}$$

Again, formal expressions can be derived for both $\overline{u}_{i,\theta,b_i}$ and $\widehat{x^u_{i,0}}(\phi, b_i)$, but they present no interest for the sake of explanation and are left in Appendix A.

Remark 1 At this stage, we acknowledge the existing similarities between our approach and the one presented in Wang et al. (2014), an extension of Ramsay et al. (2007) to a population framework. Both methods approximate the original ODE and avoid initial condition estimation. However, Wang et al. (2014) still consider classic likelihood for ϕ estimation and the absence of a proper probabilistic framework for handling $\left\{x^u_{i,0}\right\}_{i=1,\ldots n}$ makes it difficult to incorporate a priori information when available. Moreover, the spline basis decomposition used by Wang et al. (2014) is a source of inaccuracy for ODE solution reconstruction, a cause of estimation error as pointed out in Clairon (2020) in which Ramsay et al. (2007) and control based approaches have been compared in a one subject setting. Finally, the estimation quality of the method proposed in Ramsay et al. (2007) critically depends on hyperparameter choices (basis dimension, knots location etc.) which can be complicated even when data are coming from one subject and can thus become intractable for large populations.

2.2 Middle criterion

To construct an estimator $\widehat{b_i}$ of the random effects, we rely on an approximation of $\ln \mathbb{P}(\mathbf{y_i}, b_i, x^u_{i,0} \mid \phi)$ profiled on the unknown initial conditions. Since

$$\begin{aligned}\mathbb{P}(\mathbf{y_i}, b_i, x^u_{i,0} \mid \phi) &= \mathbb{P}(\mathbf{y_i} \mid \phi, b_i, x^u_{i,0})\mathbb{P}(b_i, x^u_{i,0} \mid \phi)\\ &= \mathbb{P}(\mathbf{y_i} \mid \phi, b_i, x^u_{i,0})\mathbb{P}(x^u_{i,0} \mid \phi, b_i)\mathbb{P}(b_i \mid \phi),\end{aligned}$$

with $\mathbb{P}(b_i \mid \phi) = \frac{1}{\sqrt{(2\pi)^q|\sigma^2(\triangle^T\triangle)^{-1}|}} e^{-\frac{1}{2}b_i^T\frac{\triangle^T\triangle}{\sigma^2}b_i}$, we can define as estimator:

$$\begin{aligned}\widehat{b_i}(\phi) &= \arg\min_{b_i} \min_{x^u_{i,0}} -2\ln \mathbb{P}(\mathbf{y_i}, b_i, x^u_{i,0} \mid \phi) \\ &= \arg\min_{b_i} \left\{ \min_{x^u_{i,0}} \left\{ \frac{1}{\sigma^2} \sum_j \left\| CX_{\theta,b_i,x_{i,0}}(t_{ij}) - y_{ij} \right\|_2^2 - 2\ln \mathbb{P}\left(x^u_{i,0} \mid \phi, b_i \right) \right\} \right. \\ &\quad \left. + \frac{\|\Delta b_i\|_2^2}{\sigma^2} \right\}.\end{aligned}$$

Still, we use the same relaxation & penalization scheme as in the previous section to account for model error presence for b_i^* estimation. We replace again the term $\sum_j \left\| CX_{\theta,b_i,x_{i,0}}(t_{ij}) - y_{ij} \right\|_2^2$ by $\min_{u_i} \mathcal{C}_i(x^u_{i,0}, u_i \mid \theta, b_i, U)$ in the previous criteria and we end up with the following estimator:

$$\widehat{b_i}(\phi) := \arg\min_{b_i} G_i(b_i \mid \phi) \tag{6}$$

where:

$$G_i(b_i \mid \phi) = H_i(\widehat{x^u_{i,0}}(\phi, b_i) \mid \phi, b_i) + \frac{\|\Delta b_i\|_2^2}{\sigma^2}. \tag{7}$$

2.3 Outer criterion

2.3.1 *F* general expression

We focus in this section on population parameter estimation. Classic maximum likelihood based approaches generally consider as estimator: $\widehat{\phi} := \arg\max_\phi \mathbb{P}(\phi \mid \mathbf{y}) = \arg\max_\phi \prod_i \int \mathbb{P}(\phi, b_i, x^u_{i,0} \mid \mathbf{y_i}) d(b_i, x^u_{i,0})$. This generally requires the numerical approximation of integrals of possibly high dimensions, a source of approximation and computational issues (Pinheiro and Bates 1994). To avoid this, we consider the random effects as nuisance parameters and rely on a classic profiling approach for ϕ^* estimation (Murphy and der Vaart 2000). Instead of taking the mean, we rely on the profiled joint distribution sequentially with respect to $b := \left\{ b_i \right\}_{i=1,\dots n}$ and $x^u_0 = \left\{ x^u_{i,0} \right\}_{i=1,\dots n}$, or equivalently $\min_b \min_{x^u_0} \left(-2\ln \mathbb{P}(\phi, b, x^u_0 \mid \mathbf{y}) \right)$. Bayes formula gives us $\mathbb{P}(\phi, b, x^u_0 \mid \mathbf{y}) \propto \mathbb{P}(\mathbf{y}, b, x^u_0 \mid \phi)\mathbb{P}(\phi)$ and we get $\mathbb{P}(\phi, b, x^u_0 \mid \mathbf{y}) \propto \left(\prod_i \mathbb{P}(\mathbf{y_i}, b_i, x^u_{i,0} \mid \phi) \right) \mathbb{P}(\phi)$ by conditional independence of subject by subject observations and subject specific parameters. It follows that

$$\min_b \min_{x^u_0} \left(-2\ln \mathbb{P}\left(\phi, b, x^u_0 \mid \mathbf{y}\right) \right) \propto \sum_i \min_{b_i} \min_{x^u_{i,0}} \left\{ -2\ln \mathbb{P}(\mathbf{y_i}, b_i, x^u_{i,0} \mid \phi) \right\} - 2\ln \mathbb{P}(\phi),$$

from which we derive the estimator

$$\begin{aligned}\widehat{\phi} \;=\; & \arg\min_{\phi}\left\{\sum_i \min_{b_i}\left\{\min_{x^u_{i,0}}\left\{\frac{1}{\sigma^2}\sum_j\left\|CX_{\theta,b_i,x_{i,0}}(t_{ij})-y_{ij}\right\|_2^2-2\ln\mathbb{P}(x^u_{i,0}\mid\phi,b_i)\right\}\right.\right.\\ &\left.\left.+\;\frac{\|\Delta b_i\|_2^2}{\sigma^2}\right\}\right.\\ &+\;\left.\left(d^o\sum_i n_i+nq\right)\ln\sigma^2-n\ln|\triangle^T\triangle|-2\ln\mathbb{P}(\phi)\right\}\end{aligned}$$

by using the exact expression of $\ln\mathbb{P}(\mathbf{y_i}, b_i, x^u_{i,0}\mid\phi)$ (computational details are recalled in Appendix B). In order to account for the presence of model error and limit its effect on estimation, we replace in the last expression the classic profiled likelihood estimator for b_i^* and $x^{u*}_{i,0}$ by $\widehat{b}_i(\phi)$ and $\widehat{x^u_{i,0}}(\phi,b_i)$ respectively and $X_{\theta,b_i,x_{i,0}}$ by $\overline{X}_{\theta,b_i}$. This leads us to the following population parameter estimator:

$$\widehat{\phi}\;:=\;\arg\min_{\phi}F(\phi) \tag{8}$$

where:

$$\begin{aligned}F(\phi)\;\;=\;&\frac{1}{\sigma^2}\sum_i\left(\sum_j\left\|C\overline{X}_{\theta,\widehat{b}_i(\phi)}(t_{ij})-y_{ij}\right\|_2^2+\left\|\Delta\widehat{b}_i(\phi)\right\|_2^2\right)\\ &-2\ln\mathbb{P}(\widehat{x^u_{i,0}}(\phi,\widehat{b}_i(\phi))\mid\phi,\widehat{b}_i(\phi))\\ &+\left(d^o\sum_i n_i+nq\right)\ln\sigma^2-n\ln|\triangle^T\triangle|-2\ln\mathbb{P}(\phi).\end{aligned} \tag{9}$$

2.3.2 *F* profiling on σ for uniform $x^u_{i,0}$ distribution

If $\mathbb{P}(x^u_{i,0}\mid\phi,b_i)$ is constant then $\widehat{x^u_{i,0}}(\phi,b_i)$ and $\widehat{b}_i(\phi)$ do not depend on σ i.e. $\widehat{x^u_{i,0}}(\phi,b_i)=\widehat{x^u_{i,0}}(\theta,b_i)$ and $\widehat{b}_i(\phi)=\widehat{b}_i(\theta,\Delta)$ and consequentially neither does $\overline{X}_{\theta,\widehat{b}_i(\theta,\Delta)}$. So, for each (θ,Δ), the maximizer in σ^2 of $F(\phi)=F(\theta,\Delta,\sigma)$ has a closed form expression:

$$\sigma^2(\theta,\Delta)=\frac{1}{\left(d^o\sum_i n_i+qn\right)}\sum_i\left(\sum_j\left\|C\overline{X}_{\theta,\widehat{b}_i(\theta,\Delta)}(t_{ij})-y_{ij}\right\|_2^2+\left\|\Delta\widehat{b}_i(\theta,\Delta)\right\|_2^2\right). \tag{10}$$

By using $\sigma^2(\theta,\Delta)$ expression, we get $\min_{\sigma^2}F(\phi)=\overline{F}[\theta,\Delta]$ where:

$$\overline{F}[\theta,\Delta]=\left(d^o\sum_i n_i+qn\right)\ln\left(\sigma^2(\theta,\Delta)\right)-n\ln|\triangle^T\triangle|-2\ln\mathbb{P}(\phi).$$

Thus, we can profile $F(\phi)$ on σ^2 and define our estimator as:

$$\left(\widehat{\theta},\widehat{\Delta}\right)=\;\arg\min_{(\theta,\Delta)}\overline{F}[\theta,\Delta]\,. \tag{11}$$

An estimator of σ^* is obtained from there by computing $\sigma^2\left(\widehat{\theta},\widehat{\Delta}\right)$, given by equation (10). The details of $\overline{F}$ derivation are left in appendix B.

2.4 Asymptotic Variance-Covariance matrix estimator for $\left(\widehat{\theta}, \widehat{\Delta}\right)$

We derive an estimator of the asymptotic variance of $\left(\widehat{\theta}, \widehat{\Delta}\right)$. Here we restrict to the case described in Sect. 2.3.2 when a uniform distribution is chosen for $x_{i,0}^{u}$ and the outer criterion is profiled on σ. The general case can be considered similarly, but we withdraw it for the sake of clarity since it is not used in following simulation works. We highlight that in practice the matrix Δ is parametrized by a vector δ of dimension q', i.e $\triangle := \triangle(\delta)$ and we give here a variance estimator of $\left(\widehat{\theta}, \widehat{\delta}\right)$. From this, the variance of $\widehat{\Delta}$ can be obtained using classic delta-methods (see van der Vaart (1998) chapter 3).

Let us start by presenting sufficient conditions ensuring our estimator is asymptotically normal, by introducing $h(b_i, \theta, \Delta, \mathbf{y_i}) = \left\|\Delta b_i\right\|_2^2 + \sum_j \left\|C\overline{X}_{\theta,b_i}(t_{ij}) - y_{ij}\right\|_2^2$:

1. the function $\widetilde{F}[\theta, \Delta(\delta)] = -0.5\left(d^o\mathbb{E}\left[n_1\right] + q\right)\ln\left(\frac{\lim_n \frac{1}{n}\sum_i^n \mathbb{E}\left[h(\widehat{b}_i(\theta,\Delta(\delta)),\theta,\Delta(\delta),y_i)\right]}{d^o\mathbb{E}\left[n_1\right]+q}\right) + \ln\left|\Delta(\delta)\right|$ has a well separated minimum $\left(\overline{\theta}, \overline{\delta}\right)$ belonging to the interior of a compact $\Theta \times \Omega$,
2. the true densities of unknown initial conditions $\left\{\mathbb{P}^*(.\mid \phi^*, b_i^*)\right\}_{i=1,\dots,n}$ have finite variance and either
   (a) are identicals: $\mathbb{P}^*(.\mid \phi^*, b_i^*) := \mathbb{P}^*(.\mid \phi^*)$,
   (b) or are such that for every $\varepsilon > 0$, we got
   $$\lim_{n\longrightarrow\infty} \frac{1}{\left(V^{(\nu)}\right)^2}\mathbb{E}\left[\sum_{i=1}^{n}\left(h^{(\nu)}(\mathbf{y_i}) - \mathbb{E}\left[h^{(\nu)}(\mathbf{y_i})\right]\right)^2 1_{\left\{h(\mathbf{y_i})-\mathbb{E}\left[h(\mathbf{y_i})\right]>\varepsilon\sqrt{V^{(\nu)}}\right\}}\right] = 0$$
   for $\nu = 0, 1$ where $h^{(\nu)}(\mathbf{y_i}) = \frac{d^{(\nu)}h}{d^{(\nu)}(\theta,\delta)}(\widehat{b}_i(\overline{\theta}, \Delta(\overline{\delta})), \overline{\theta}, \Delta(\overline{\delta}), \mathbf{y_i})$ and $V^{(\nu)} = \sqrt{\sum_i Var(h^{(\nu)}(\mathbf{y_i}))^2}$,
3. the subject specific number of observations $\left\{n_i\right\}_{i=1,\dots,n}$ are i.i.d and uniformly bounded,
4. for all possibles values $\left(\theta, b_i\right)$, the solution $X_{\theta,b_i,x_{i,0}^*}$ belongs to a compact χ of $\mathbb{R}^d$, and for all (t, θ, x), the mapping $b_i \longmapsto f_{\theta,b_i}(t, x)$ has a compact support Θ_b,
5. $\left(\theta, b_i, t, x\right) \longmapsto f_{\theta,b_i}(t, x)$ belongs to $C^1(\Theta \times \Theta_b \times [0,\ T] \times \chi, \mathbb{R}^d)$,
6. the matrices $\frac{\partial^2 \mathcal{C}_i}{\partial^2 x_{i,0}}(\widehat{x_{i,0}^u}(\theta, \Delta(\overline{\delta})), \overline{u}_{\overline{\theta},\widehat{b}_i(\overline{\theta},\Delta(\overline{\delta}))} \mid \overline{\theta}, \widehat{b}_i\left(\overline{\theta}, \Delta(\overline{\delta})\right), U)$ and $\frac{\partial^2}{\partial^2 b_i}G_i(\widehat{b}_i\left(\overline{\theta}, \Delta(\overline{\delta})\right) \mid \overline{\theta}, \Delta(\overline{\delta}))$ are of full rank almost surely for every sequence $\mathbf{y_i}$,
7. there is a neighborhood $\Theta_{\overline{\theta}}$ of $\overline{\theta}$ such that $\left(\theta, b_i, t, x\right) \longmapsto f_{\theta,b_i}(t, x) \in C^5(\Theta_{\overline{\theta}} \times \Theta_b \times [0,\ T] \times \chi, \mathbb{R}^d)$.

Condition 2b is here to ensure asymptotic normality for non identically distributed random variables via Lindeberg-Feller theorem. Conditions 1–4 are

used to derive the consistency of our estimator toward $\left(\overline{\theta},\overline{\delta}\right)$ by following classic steps for M-estimator by proving 1. the uniform convergence of our stochastic cost function to a deterministic one, 2. the existence of a well-separated minimum for this deterministic function (van der Vaart 1998). Conditions 5–7 ensures that our cost function is asymptotically smooth enough in the vicinity of $\left(\overline{\theta},\overline{\delta}\right)$ to proceed to a Taylor expansion and transfer the regularity of the cost function to the asymptotic behavior of $\sqrt{n}(\widehat{\theta}-\overline{\theta},\widehat{\delta}-\overline{\delta})$. Less restrictive conditions can be established under which our estimator is still asymptotically normal, in particular regarding f_{θ,b_i} regularity with respect to t.

Theorem 1 *Under conditions* 1–7, *there is a model dependent lower bound* λ *such that if* $\|U\|_2 > \lambda$ *then the estimator* $\left(\widehat{\theta},\widehat{\delta}\right)$ *converges almost surely to* $\left(\overline{\theta},\overline{\delta}\right)$ *such that*:

$$\sqrt{n}(\widehat{\theta}-\overline{\theta},\widehat{\delta}-\overline{\delta}) \rightsquigarrow N\left(0, A(\overline{\theta},\overline{\delta})^{-1}B(\overline{\theta},\overline{\delta})\left(A(\overline{\theta},\overline{\delta})^{-1}\right)^T\right)$$

where

$$A(\overline{\theta},\overline{\delta}) = \lim_n \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\partial J(\overline{\theta},\overline{\delta},\mathbf{y_i})}{\partial(\theta,\delta)}\right], B(\overline{\theta},\overline{\delta}) = \lim_n \frac{1}{n}\left[\sum_i J(\overline{\theta},\overline{\delta},\mathbf{y_i})J(\overline{\theta},\overline{\delta},\mathbf{y_i})^T\right]$$

and the vector valued function $J(\theta,\delta,\mathbf{y_i}) = \begin{pmatrix} J_\theta(\theta,\delta,\mathbf{y_i}) \\ J_\delta(\theta,\delta,\mathbf{y_i}) \end{pmatrix}$ *is given by*:

$$\begin{aligned} J_\theta(\theta,\delta,\mathbf{y_i}) &= \frac{d}{d\theta}h(\hat{b}(\theta,\Delta(\delta)),\theta,\Delta(\delta),y_i) \\ J_\delta(\theta,\delta,\mathbf{y_i}) &= \frac{d}{d\delta}h(\hat{b}_i(\theta,\Delta(\delta)),\theta,\Delta(\delta),y_i) \\ &\quad - \frac{2}{d^o\mathbb{E}\left[n_1\right]+q}Tr\left(\vartriangle(\delta)^{-1}\frac{\partial\vartriangle(\delta)}{\partial\delta_k}\right)h(\hat{b}_i(\theta,\Delta(\delta)),\theta,\Delta(\delta),y_i). \end{aligned}$$

The proof is left in appendix D. The practical interest of this theorem is to give an estimator of the Variance-Covariance matrix: $V(\widehat{\theta},\widehat{\delta}) \simeq \widehat{A}(\widehat{\theta},\widehat{\delta})^{-1}\widehat{B}(\widehat{\theta},\widehat{\delta})\left(\widehat{A}(\widehat{\theta},\widehat{\delta})^{-1}\right)^T/n$ with $\widehat{A}(\widehat{\theta},\widehat{\delta}) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\widehat{J}(\widehat{\theta},\widehat{\delta},\mathbf{y_i})}{\partial(\theta,\delta)}$, $\widehat{B}(\widehat{\theta},\widehat{\delta}) = \frac{1}{n}\sum_{i=1}^{n}\widehat{J}(\widehat{\theta},\widehat{\delta},\mathbf{y_i})\widehat{J}(\widehat{\theta},\widehat{\delta},\mathbf{y_i})^T$ and the vector valued function $\widehat{J}(\theta,\delta,\mathbf{y_i}) = \begin{pmatrix} \widehat{J}_\theta(\theta,\delta,\mathbf{y_i}) \\ \widehat{J}_\delta(\theta,\delta,\mathbf{y_i}) \end{pmatrix}$ given by $\widehat{J}_\theta(\theta,\delta,\mathbf{y_i}) = J_\theta(\theta,\delta,\mathbf{y_i})$ and

$$\begin{aligned} \widehat{J}_\delta(\theta,\delta,\mathbf{y_i}) &= \frac{d}{d\delta}h(\widehat{b}_i(\theta,\Delta(\delta)),\theta,\Delta(\delta),\mathbf{y_i}) \\ &\quad - \frac{2n}{d^o\sum_i n_i+qn}Tr\left(\bigtriangleup(\delta)^{-1}\frac{\partial\bigtriangleup(\delta)}{\partial\delta_k}\right)h(\widehat{b}_i(\theta,\Delta(\delta)),\theta,\Delta(\delta),\mathbf{y_i}). \end{aligned}$$

Now that we have proven the existence of the variance-covariance matrix $V(\overline{\theta}, \overline{\delta})$ such that $\widehat{\delta} - \overline{\delta} \rightsquigarrow N\left(0, V(\overline{\theta}, \overline{\delta})\right)$, we can use the Delta method to derive the asymptotic normality of the original matrix $\Psi\left(\widehat{\delta}\right) = \sigma^2\left(\Delta(\widehat{\delta})^T\Delta(\widehat{\delta})\right)^{-1}$ as well as an estimator of its asymptotic variance. In the case of a diagonal matrix Ψ, composed of the elements $\left(\Psi_1^2, \dots \Psi_q^2\right)$ and of the parametrization $\triangle(\delta) = Diag(\left\{e^{\delta_l}\right\})_{l=1,\dots q'}$ used in Sect. 3, we derive:

$$\begin{pmatrix} \Psi_1(\widehat{\delta}) \\ \vdots \\ \Psi_q(\widehat{\delta}) \end{pmatrix} - \begin{pmatrix} \Psi_1(\delta^*) \\ \vdots \\ \Psi_q(\delta^*) \end{pmatrix} \rightsquigarrow N\left(0, \sigma^2 \begin{pmatrix} e^{-\delta_1^*} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & e^{-\delta_q^*} \end{pmatrix} V(\theta^*, \delta^*) \begin{pmatrix} e^{-\delta_1^*} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & e^{-\delta_q^*} \end{pmatrix}\right).$$

Remark 2 The previous theorem 1 states that we retrieve a parametric convergence rate. Thus, we avoid the pitfall described in Sartori (2003) for profiled methods in presence of a number of nuisance parameters increasing with the number of subjects (or strata to resume Sartori (2003) terminology) potentially leading to bias accumulation for score functions among subjects. The i.i.d structure of random effects allows us to rely on central limit theorem to avoid this accumulation phenomenon.

3 Results on simulated data

We compare the accuracy of our approach with maximum likelihood (ML) in different models and experimental designs reflecting the problems exposed in the introduction, that is estimation in 1. presence of model error, 2. partially observed framework with unknown initial conditions and 3. presence of poorly identifiable parameters. We proceed to Monte-Carlo simulations based on $N_{MC} = 100$ runs. At each run, we generate n_i observations coming from n subjects on an observation interval $[0, T]$ with Gaussian measurement noise of standard deviation σ^*. From these data, we estimate θ^*, Ψ^* and b_i^* with both estimation methods. We quantify the accuracy of each estimator $\widehat{\psi}_p$ of the population parameters estimate $\widehat{\psi} = \left(\widehat{\theta}, \widehat{\Psi}\right)$ via Monte-Carlo computation of the bias $Bias(\widehat{\psi}_p) = \mathbb{E}\left[\widehat{\psi}_p\right] - \psi_p^*$, the empirical variance $V^e(\widehat{\psi}_p) = \mathbb{E}\left[\left(\mathbb{E}\left[\widehat{\psi}_p\right] - \psi_p^*\right)^2\right]$, the mean squared error $MSE(\widehat{\psi}_p) = Bias(\widehat{\psi}_p)^2 + V^e(\widehat{\psi}_p)$, the estimated variance $\widehat{V}\left(\widehat{\psi}_p\right)$, as well as the coverage rate of the 95%-confidence interval derived from it. This coverage rate, denoted CR in the following results, corresponds to the frequency at which the interval $\left[\widehat{\psi}_p \pm z_{0.975}\sqrt{\widehat{V}\left(\widehat{\psi}_p\right)}\right]$ contains ψ_p^* with $z_{0.975}$ the 0.975−quantile of the centered Gaussian law. We compute the previous quantities for the normalized values $\widehat{\psi}_p^{norm} := \frac{\widehat{\psi}_p}{\psi_p^*}$ to make relevant comparisons among parameters with different order of

magnitude. For b_i^*, we estimate the mean squared error $MSE(\widehat{b}_i) = \mathbb{E}\left[\left\|b_i^* - \widehat{b}_i\right\|_2^2\right]$. For each subsequent examples, we give the results for $n = 50$ and present in Appendix C the case $n = 20$ to analyze the evolution of each estimator accuracy with respect to data sparsity. For one example in section "Effect of population size on estimation accuracy" in Appendix C, we also consider the case $n = 100$ to analyze the evolution of estimation accuracy with respect to an increasing population size.

In the following, we use the superscript *ML* to denote the ML estimator. For the fairness of comparison with ML, we choose a non-informative prior i.e. $\ln \mathbb{P}(\theta, \Delta) = 0$ for our method throughout this section (the impact of prior incorporation is analyzed in an example left in Appendix C, section "Effect of prior information on estimation accuracy", for a discussion on prior choice in NLME-ODEs see Prague et al. (2012)). Also, we do not use a distribution for $x_{i,0}^u$ for our approach. For ML which requires it, we will use the right parametric form for $\mathbb{P}\left(x_{i,0}^u \mid \phi, b_i\right)$. If the ODE (1) has an analytical solution, the ML estimator is computed via SAEM algorithm (SAEMIX package Comets et al. (2017)). Otherwise, it is done via a restricted likelihood method dedicated to ODE models implemented in the nlmeODE package (Tornoe et al. 2004). For our method, we need to select U balancing model and data fidelity in the inner and middle criteria (5)-(7). We use the method presented in G. Hooker and Earn (2011) to compute $\mathrm{EP}_i(U)$, the prediction error for the subject i corresponding to the estimators $\widehat{\theta_U}$, $\left\{\widehat{b_{i,U}}\right\}_{i=1,\dots n}$ obtained for a given matrix U. From this, we compute $\mathrm{EP}(U) = \sum_i \mathrm{EP}_i(U)$ the global prediction error for the whole population. We test a trial of scalar matrices $\left\{U_l\right\}_{l=1,\dots L} = \left\{\lambda_l \times I_d\right\}_{l=1,\dots L}$ and retain the hyperparameter value λ_l minimizing EP and we denote $\widehat{\theta}, \widehat{\Psi}$, $\left\{\widehat{b_{i,}}\right\}_{i=1,\dots n}$ the corresponding estimator. For solving the optimization problems required for computing our criteria, we use the Nelder-Mead algorithm implemented in the optimr package (Nash 2016). All optimization algorithms used here require a starting guess value. We start from the true parameter value for each of them. By doing so, we aim to keep distinct two problems: 1. the numerical stability of the estimation procedures, 2. the intrinsic accuracy of the different estimators. These two problems are correlated, but we aim to address only the latter which corresponds to the issues raised in introduction. Still, we check on preliminary analysis that local minima presence was not an issue in the neighborhood of $\left(\theta^*, \triangle^*\right)$ by testing different starting points for all methods. No problem appears for our method and SAEMIX. A negligible number of non convergence cases appear for nlmeODE which have been discarded thanks to the convergence criteria embedded in the package (the occurence and importance of such convergence issues is analyzed in an example left in section "Effect of wrong first guess on estimation accuracy" in Appendix C in which we show that our method suffers less than ML from convergence issues when initial conditions are unknown).

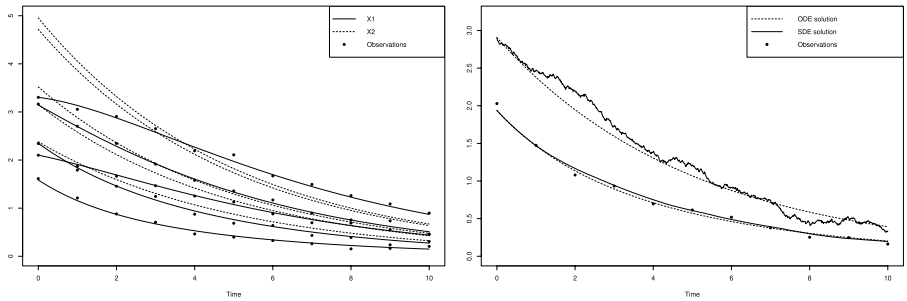


Fig. 1 Left: Examples of solutions of (12) and corresponding observations. Right: Solution of (12) and a realization of (13) for the same parameter values

3.1 Application 1—partially observed linear model

We consider the population model where each subject i follows the ODE:

Table 1 Results of estimation for model (12). The different subscripts stand for the following estimation scenarios: 1. x_0 when both initial conditions are set to $\left(x^*_{1,i,0}, x^*_{2,i,0}\right)$, 2. $x_{0,2}$ when $x_{1,i,0}$ is set to $y_{i,0}$ and $x_{2,i,0}$ to $x^*_{2,i,0}$, 3. absence of subscript when $x_{1,i,0}$ is set to $y_{i,0}$ and $x_{2,i,0}$ is estimated. Results from our method are in bold

		Well-specified						Misspecified					
		MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i	MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i
θ_1	$\widehat{\theta}^{ML}_{x_0}$	0.01	3e−3	0.01	0.01	0.95		0.02	3e−3	0.01	0.01	0.92	
	$\widehat{\theta}^{ML}_{x_{2,0}}$	0.01	3e−3	0.01	0.01	0.92		0.01	3e−3	0.01	0.01	0.93	
	$\widehat{\theta}^{ML}$	0.02	4e−3	0.02	0.01	0.90		0.02	0.02	0.02	0.01	0.90	
	$\widehat{\theta}$	**0.01**	**6e−3**	**7e−3**	**0.01**	**0.98**		**0.01**	**0.01**	**0.01**	**0.01**	**0.96**	
θ_2	$\widehat{\theta}^{ML}_{x_0}$	2e−4	1e−3	1e−4	1e−4	0.95		1e−3	0.02	4e−4	6e−4	0.88	
	$\widehat{\theta}^{ML}_{x_{2,0}}$	2e−4	1e−3	1e−4	1e−4	0.93		1e−3	0.02	5e−4	6e−4	0.88	
	$\widehat{\theta}^{ML}$	5e−4	4e−3	5e−4	3e−4	0.86		5e−3	0.02	4e−3	1e−3	0.77	
	$\widehat{\theta}$	**4e−4**	**3e−4**	**4e−4**	**5e−4**	**0.98**		**3e−3**	**0.01**	**3e−3**	**5e−3**	**0.94**	
Ψ	$\widehat{\theta}^{ML}_{x_0}$	0.01	−0.02	0.01	0.01	0.93	6e−3	0.01	−0.02	0.01	0.01	0.93	0.01
	$\widehat{\theta}^{ML}_{x_{2,0}}$	0.01	−0.02	0.01	0.01	0.92	7e−3	0.01	−0.02	0.01	0.01	0.94	0.02
	$\widehat{\theta}^{ML}$	0.01	−0.02	0.01	0.01	0.92	0.01	0.01	−0.02	0.01	0.01	0.93	0.04
	$\widehat{\theta}$	**0.01**	**− 0.02**	**0.01**	**0.01**	**0.92**	**5e−3**	**0.01**	**− 0.03**	**0.01**	**0.02**	**0.92**	**0.02**

$$\begin{cases} \dot{X}_{1,i} = \phi_{2,i} X_{2,i} - \phi_{1,i} X_{1,i} \\ \dot{X}_{2,i} = -\phi_{2,i} X_{2,i} \\ \left(X_{1,i}(0), X_{2,i}(0)\right) = \left(x_{1,i,0}, x_{2,i,0}\right) \end{cases} \tag{12}$$

with the following parametrization: $\log(\phi_{1,i}) = \theta_1 + b_i$ and $\log(\phi_{2,i}) = \theta_2$ where $b_i \sim N(0, \Psi)$. The true population parameter values are $\theta^* = (\theta_1^*, \theta_2^*) = (\log(0.5), \log(2))$ and $\Psi^* = 0.5^2$ and we are in a partially observed framework where only $X_{1,i}$ is accessible. The true initial conditions are distributed with $x_{1,i,0}^* \sim N(2,\ 0.5)$ and $x_{2,i,0}^* \sim N(3,\ 1)$. For the penalization term in our method, we choose the values $\lambda_l = 10^4, 10^6, 10^8$. An analytic solution exists for ODE (12). In particular the first component is given by $X_{1,i}(t) = e^{-\phi_{1,i}t}(x_{1,i,0} + \frac{x_{2,i,0}\phi_{2,i}}{\phi_{1,i}-\phi_{2,i}}(e^{(\phi_{1,i}-\phi_{2,i})t} - 1))$ and will be used for estimation with the SAEMIX package. We generate $n_i = 11$ longitudinal observations per subject on $[0,\ T] = [0,\ 10]$ with measurement noise of standard deviation $\sigma = 0.03$. An example of sampled observations and corresponding solutions are plotted in Fig. 1.

We want to investigate the impact of initial condition, especially the unobserved one $x_{2,i,0}^*$, on the ML estimator accuracy. Indeed, our method does not need to estimate $x_{2,i,0}^*$ and thus no additional difficulties appear in this partially observed framework. For the ML, however, it is a nuisance subject-specific parameter that should be estimated and for which no observations are available. For this, we compute $\widehat{\theta}^{ML}_{x_0}$, $\widehat{\theta}^{ML}_{x_{2,0}}$ and $\widehat{\theta}^{ML}$ the ML estimator respectively when: 1. both initial conditions are perfectly known, 2. $x_{1,i,0}^*$ is replaced by the measured value, 3. in addition, $x_{2,i,0}^*$ has to be estimated.

3.1.1 Well-specified case

We used the exact model described in Sect. 3.1 for the estimation procedure. Thus, we are in a completely well-specified setting, with all mechanisms modeled. We present the estimation results in Table 1—left side. For ML, the results are goods in terms of accuracy and consistent in terms of asymptotic confidence interval coverage rate when both initial conditions are known: 95% for θ_1 and θ_2, which is consistent with theoretical results. However, there is a significant drop in accuracy when $x_{2,i,0}^*$ has to be estimated. In particular, the coverage rate drops to 90% and 86% for θ_1 and θ_2 respectively. Interestingly, ML inaccuracy is driven by bias and under-estimated variance when initial conditions are not known (as shown by a greater V^e than $\widehat{V}$). In this case our method provides a relevant alternative: it gives accurate estimations with a good coverage rate for all parameters while avoiding the estimation of $x_{2,i,0}^*$. Variances are properly estimated compared to empirical variances. Estimation of individual random effects is also more accurate with our method, with a MSE for b_i 2 times smaller compared to ML with unknown initial conditions.

3.1.2 Misspecified case in presence of model error at the subject level

To mimic the presence of misspecification, we now generate the observations from the hypoelliptic stochastic model:

$$\begin{cases} dX_{1,i} = \phi_{2,i} X_{2,i} dt - \phi_{1,i} X_{1,i} dt \\ dX_{2,i} = -\phi_{2,i} X_{2,i} dt + \alpha dB_t \\ \left(X_{1,i}(0), X_{2,i}(0)\right) = \left(x_{1,i,0}, x_{2,i,0}\right) \end{cases} \tag{13}$$

with B_t a Wiener process and $\alpha = 0.1$ the diffusion coefficient. For the sake of comparison, a solution of (12) and a realization of its perturbed counterpart given by (13) are plotted in Fig. 1. This framework where stochasticity only affects the unmeasured compartment is known to be problematic for parameter estimation and inference procedures are yet to be developed for sparse sampling case. From Fig. 1 it is easy to see that the diffusion α will be hard to estimate when we only have observations for $X_{1,i}$. Thus, we still estimate the parameters from the model (12) which is now seen as a deterministic approximation of the true stochastic process. Still, it is expected that our method will mitigate the effect of stochasticity on the estimation accuracy by taking into account model misspecification. Results are presented in Table 1—right side. The differences between the two methods are similar to the previous well-specified case with an additional loss of accuracy coming from model error for both estimators. However, the misspecification effect for ML is more pronounced comparing to our method which manages to limit the damages done. This illustrates the benefits of taking into account model uncertainty for estimation, in particular here when model error occurs in the unobserved compartment, a situation in which classic statistical criteria for model assessment based on a data fitting criterion are difficult to use.

Finally, we acknowledge that the effect of other misspecification sources can be investigated. For example, the population which is here assumed homogeneous can be in fact a mixture of subjects with random effects distributed according to different laws. To account for this, we evaluate in section "Effect of outlier presence on estimation accuracy" in Appendix C the situation in which an increasing fraction of subjects are chosen as outliers for the random effect assumed distribution. We then investigate its impact on estimation accuracy for ML and optimal control based methods.

3.2 Application 2–Partially observed nonlinear model

We consider the model presented in De Gaetano and Arino (2000) for the analysis of glucose and insulin regulation:

Table 2 Results of estimation for model (14). The different subscripts stand for the following estimation scenarios: 1. S_I when S_I is set to S_I^*, 2. absence of subscript when S_I is estimated. Results from our method are in bold

		Well-specified						Misspecified					
		MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i	MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i
θ_{S_G}	$\widehat{\theta}^{ML}_{S_I}$	5e−5	2e−3	4e−5	9e−6	0.95		6e−5	3e−3	6e−5	2e−5	0.85	
	$\widehat{\theta}^{ML}$	2e−3	0.03	1e−3	8e−5	0.85		2e−3	3e−3	1e−3	2e−4	0.54	
	$\widehat{\theta}_{S_I}$	**1e−5**	**4e−4**	**1e−5**	**8e−6**	**0.95**		**2e−5**	**− 2e−5**	**2e−5**	**2e−5**	**0.93**	
	$\widehat{\theta}$	**2e−4**	**− 6e−4**	**2e−4**	**2e−4**	**0.96**		**3e−4**	**− 1e−3**	**3e−4**	**4e−4**	**0.93**	
θ_{S_I}	$\widehat{\theta}^{ML}_{S_I}$	known						known					
	$\widehat{\theta}^{ML}$	2e−3	0.03	1e−3	6e−5	0.90		0.01	0.04	0.01	1e−3	0.55	
	$\widehat{\theta}_{S_I}$	**known**						**known**					
	$\widehat{\theta}$	**1e−4**	**− 7e−4**	**1e−4**	**1e−4**	**0.96**		**3e−4**	**− 1e−3**	**3e−4**	**3e−4**	**0.92**	
θ_m	$\widehat{\theta}^{ML}_{S_I}$	7e−4	3e−3	6e−4	5e−4	0.94		8e−4	− 3e−3	8e−4	5e−4	0.89	
	$\widehat{\theta}^{ML}$	9e−4	8e−3	8e−4	5e−4	0.86		5e−3	− 5e−3	5e−3	5e−4	0.88	
	$\widehat{\theta}_{S_I}$	**5e−4**	**6e−3**	**5e−4**	**5e−4**	**0.95**		**4− 4**	**7e−4**	**4e−4**	**5e−4**	**0.95**	
	$\widehat{\theta}$	**6e−4**	**6e−3**	**5e−4**	**5e−4**	**0.95**		**4e−4**	**6e−4**	**4e−4**	**5e−4**	**0.96**	
Ψ	$\widehat{\theta}^{ML}_{S_I}$	0.02	7e−4	0.02	0.02	0.95	0.02	0.03	− 3e−3	0.03	0.02	0.93	0.03
	$\widehat{\theta}^{ML}$	0.04	−0.09	0.03	0.02	0.88	0.02	0.03	− 8e−3	0.02	0.02	0.87	0.03
	$\widehat{\theta}_{S_I}$	**0.01**	**− 2e−3**	**0.01**	**0.01**	**0.95**	**0.01**	**0.01**	**− 4e−3**	**0.01**	**0.02**	**0.94**	**0.01**
	$\widehat{\theta}$	**0.01**	**3e−3**	**0.01**	**0.01**	**0.94**	**0.01**	**0.02**	**− 7e−3**	**0.02**	**0.02**	**0.94**	**0.02**

$$\begin{cases} \dot{G}_i = S_G(G_B - G_i) - X_i G_i \\ \dot{I}_i = \gamma t(G_i - h) - m_i(I_i - I_B) \\ \dot{X}_i = -p_2(X_i + S_I(I_i - I_B)). \end{cases} \tag{14}$$

The ODE system (14) rules the behavior of circulating glucose G_i and insulin I_i in blood as well as insulin X_i present in interstitial fluid. We are in a partially observed case where only G_i and I_i are measured. The values of parameters $\left(p_2, \gamma, h, G_B, I_B\right)$ are set to $(-4.93, -6.85, 4.14, 100, 100)$ and we aim to estimate $\theta = \left(\theta_{S_G}, \theta_{S_I}, \theta_m\right)$, linked to the original model via the parametrization: $\log(S_G) = \theta_{S_G}$, $\log(S_I) = \theta_{S_I}$ and $\log(m_i) = \theta_m + b_i$ where $b_i \sim N(0, \Psi)$. The true population parameter values are $\theta^* = (-3.89, -7.09, -1.81)$ and $\Psi^* = 0.26^2$. The true subject-specific initial conditions $x^*_{i,0} = \left(G^*_{i,0}, I^*_{i,0}, X^*_{i,0}\right)$ are distributed according to $\ln(x^*_{i,0}) \sim N(l_{x_0^*}, \Psi_{l_{x_0^*}})$ with $l_{x_0^*} = (5.52, 4.88, -7)$ and $\Psi_{l_{x_0^*}} = \left(0.17^2,\ 0.1^2, 10^{-4}\right)$. For the penalization term in our

method, we choose the values $\lambda_l = 10^6, 10^7, 10^8$. We generate $n_i = 5$ observations on $[0,\ T] = [0,\ 180]$ with Gaussian measurement noise of standard deviation $\sigma^* = 3$. As in the previous example, we investigate the impact of unknown initial conditions on the estimators accuracy. We are particularly interested in the joint estimation of θ_{S_I}, which appears only in the equation ruling the unobserved state variable X_i, and $x^*_{i,0}$ required for each subject by ML. For this, we distinguish two cases, 1. when θ_{S_I} is known, 2. when θ_{S_I} has to be estimated and we denote respectively $\widehat{\theta_{S_I}}$ and $\widehat{\theta}$ the corresponding estimators. Finally, since the model is nonlinear, we have to specify a pseudo-linear representation of the vector field as in (4):

$$A_{\theta,b_i}\left(t, G_i, I_i, X_i\right) = \begin{pmatrix} -S_G & 0 & -G_i \\ \gamma t & -m_i & 0 \\ 0 & -p_2 S_I & -p_2 \end{pmatrix},\ r_{\theta,b_i}(t) = \begin{pmatrix} S_G G_B \\ -\gamma t h + m_i I_B \\ p_2 S_I I_B \end{pmatrix}.$$

3.2.1 Well-specified case

We present the estimation results in Table 2—left side. Our method has small bias and achieve good coverages in all cases. We obtain smaller MSE than ML and avoid the drop in coverage rate of the confidence interval in the case of $\theta^*_{S_I}$ estimation, which is often needed in practice. The difference between the two estimators behavior is explained by the fact that they are defined through the construction of two different optimization problems. At the population level, our approach leads to minimize a cost function depending on a 4-dimensional parameter whereas ML, due to its need to estimate $x^*_{i,0}$, considers a 10-dimensional one. Thus, the parameter spaces explored by each method to look for the minimum are very different.

Table 3 Biological interpretation and parameter values

Parameters		Biological interpretation	Values
δ_L		Long-lived B-cells declining rate	$\log(2)/(364 \times 6)$
θ^*	$\theta^*_{\delta_S}$	Mean log-value for δ_S, the short-lived cells declining rate	$\log(\log(2)/1.2) \simeq -0.54$
	$\theta^*_{\phi_S}$	Mean log-value for ϕ_S, the antibodies influx from short-lived cells	$\log(2755) \simeq 7.92$
	$\theta^*_{\phi_L}$	Mean log-value for ϕ_L, the antibodies influx from long-lived cells	$\log(16) \simeq 2.78$
	$\theta^*_{\delta_{Ab}}$	Mean log-value for δ_{Ab}, the antibodies declining rate	$\log(\log(2)/24) \simeq -3.54$
Ψ^*	$\Psi^*_{\phi_S}$	Inter individual variance for $\log(\phi_{S,i})$	0.92^2
	$\Psi^*_{\phi_L}$	Inter individual variance for $\log(\phi_{L,i})$	0.85^2
	$\Psi^*_{\delta_{Ab}}$	Inter individual variance for $\log(\delta_{Ab,i})$	0.3^2

3.2.2 Misspecified case in presence of model error at the subject level

To mimic misspecification presence, we generate the observations from the stochastic model:

$$\begin{cases} dG_i = \big(S_G(G_B - G_i) - X_i G_i\big)dt + \alpha_1 dB_{1,t} \\ dI_i = \big(\gamma t(G_i - h) - m_i(I_i - I_B)\big)dt + \alpha_2 dB_{2,t} \\ dX_i = \big(-p_2(X_i + S_I(I_i - I_B))\big)dt + \alpha_3 dB_{3,t} \end{cases}, \tag{15}$$

where the $B_{i,t}$ are Wiener processes and $\big(\alpha_1, \alpha_2, \alpha_3\big) = \big(2, 2, 2 \times 10^{-4}\big)$ their diffusion coefficients. We present the estimation results in Table 2—right side. For ML, the drop in coverage rate for $\theta^*_{S_G}$ and $\theta^*_{S_I}$ is even more striking when $\theta^*_{S_I}$ needs to be estimated. This is explained by the effect of model misspecification which increases bias and the fact that ML does not take into account this new source of uncertainty which leads to under-estimation of variance and too narrow confidence intervals. Our method achieved small bias, nominal coverages and small MSE for random effects.

3.3 Application 3—antibody concentration evolution model

We consider the model presented in Pasin et al. (2019) to analyze the antibody concentration, denoted A_i, generated by two populations of antibody secreting cells: the short lived, denoted S_i, and the long-lived, denoted L_i:

$$\begin{cases} \dot{S}_i = -\delta_S S_i \\ \dot{L}_i = -\delta_L L_i \\ \dot{A}_i = \vartheta_{S,i} S_i + \vartheta_{L,i} L_i - \delta_{Ab,i} A_i \\ \big(S_i(0), L_i(0), A_i(0)\big) = \big(S_{i,0}, L_{i,0}, A_{i,0}\big). \end{cases} \tag{16}$$

This model is used to quantify the humoral response on different populations after an Ebola vaccine injection with a 2 doses regimen seven days after the second injection when the antibody secreting cells enter in a decreasing phase. These cells being unobserved, the preceding equation can be simplified to focus on antibody concentration evolution:

$$\dot{A}_i = \phi_{S,i} e^{-\delta_S t} + \phi_{L,i} e^{-\delta_L t} - \delta_{Ab,i} A_i \tag{17}$$

with $\phi_{S,i} := \vartheta_{S,i} S_{i,0}$ and $\phi_{L,i} := \vartheta_{L,i} L_{i,0}$. This equation has an analytic solution which will be used for ML. We consider the following parametrization: $\log(\delta_S) = \theta_{\delta_S}$, $\log(\phi_{S,i}) = \theta_{\phi_S} + b_{\phi_S,i}$, $\log(\phi_{L,i}) = \theta_{\phi_L} + b_{\phi_L,i}$ and $\log(\delta_{Ab,i}) = \theta_{\delta_{Ab}} + b_{\delta_{Ab},i}$. The true parameter values are presented in Table 3. For the penalization term in our method, we choose the values $\lambda_l = 10^3, 10^5, 10^7, 10^8$.

According to Pasin et al. (2019), δ_L was non-identifiable based on the available data and only a lower bound has been derived for it via profiled likelihood. So, to make fair comparisons between our approach and maximum likelihood, we do not

Table 4 Results of estimation for model (17). The different subscripts stand for the following estimation scenarios: 1. δ_S when θ_{δ_S} is set to $\theta^*_{\delta_S}$, 2. absence of subscript when θ_{δ_S} is estimated. Results from our method are in bold

		Well-specified						Misspecified					
		MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i	MSE	Bias	V^e	$\widehat{V}$	CR	MSE b_i
θ_{δ_S}	$\widehat{\theta}^{ML}_{\delta_S}$	known						known					
	$\widehat{\theta}_{ML}$	2.13	0.78	1.51	70.64	0.92		3.88	1.48	1.68	4.10	0.80	
	$\widehat{\theta}_{\delta_S}$	**known**						**known**					
	$\widehat{\theta}$	**0.62**	**− 0.34**	**0.50**	**0.66**	**0.92**		**0.93**	**− 0.40**	**0.77**	**0.62**	**0.90**	
θ_{ϕ_S}	$\widehat{\theta}^{ML}_{\delta_S}$	4e−4	0.01	3e−4	3e−4	0.94		1e−3	0.02	1e−3	5e−4	0.91	
	$\widehat{\theta}_{ML}$	0.01	−0.05	7e−3	0.40	0.92		0.02	−0.10	0.01	0.02	0.88	
	$\widehat{\theta}_{\delta_S}$	**2e−3**	**− 0.05**	**2e−4**	**1e−3**	**0.94**		**7e−4**	**− 0.02**	**3e−4**	**1e−3**	**0.92**	
	$\widehat{\theta}$	**2e−3**	**1e−3**	**2e−3**	**2e−3**	**0.93**		**4e−3**	**− 6e−3**	**3e−3**	**0.01**	**0.90**	
θ_{ϕ_L}	$\widehat{\theta}^{ML}_{\delta_S}$	3e−3	0.02	3e−3	2e−3	0.95		5e−3	0.03	4e−3	3e−3	0.93	
	$\widehat{\theta}_{ML}$	4e−3	0.03	4e−3	3e−3	0.90		9e−3	0.05	7e−3	4e−3	0.90	
	$\widehat{\theta}_{\delta_S}$	**7e−4**	**− 0.01**	**5e−4**	**3e−3**	**0.95**		**2e−3**	**− 0.02**	**3e−3**	**2e−3**	**0.97**	
	$\widehat{\theta}$	**3e−3**	**− 3e−3**	**3e−3**	**2e−3**	**0.91**		**6e−3**	**− 8e−3**	**6e−3**	**7e−3**	**0.90**	
$\theta_{\delta_{Ab}}$	$\widehat{\theta}^{ML}_{\delta_S}$	7e−4	− 0.02	5e−4	3e−4	0.93		2e−3	− 0.03	1e−3	1e−3	0.92	
	$\widehat{\theta}_{ML}$	2e−3	−0.02	1e−3	4e−4	0.88		4e−3	−0.04	3e−3	7e−4	0.88	
	$\widehat{\theta}_{\delta_S}$	**2e−4**	**0.01**	**1e−4**	**3e−4**	**0.95**		**3e−4**	**2e−3**	**3e−4**	**3e−4**	**0.96**	
	$\widehat{\theta}$	**4e−4**	**0.01**	**3e−4**	**2e−4**	**0.90**		**3e−4**	**8e−3**	**3e−4**	**2e−3**	**0.89**	
Ψ_{ϕ_S}	$\widehat{\theta}^{ML}_{\delta_S}$	0.04	− 1e−3	0.04	0.07	1	0.15	0.05	0.03	0.05	0.08	1	0.17
	$\widehat{\theta}_{ML}$	0.11	0.01	0.11	0.05	1	0.17	0.13	0.01	0.13	0.25	1	0.21
	$\widehat{\theta}_{\delta_S}$	**0.02**	**8e−3**	**0.02**	**0.01**	**0.94**	**0.06**	**0.02**	**2e−3**	**0.02**	**0.02**	**0.94**	**0.11**
	$\widehat{\theta}$	**0.02**	**− 0.03**	**0.02**	**0.02**	**0.94**	**0.07**	**0.02**	**− 0.05**	**0.02**	**0.03**	**0.92**	**0.08**
Ψ_{ϕ_L}	$\widehat{\theta}^{ML}_{\delta_S}$	0.03	0.04	0.02	0.04	1	0.30	0.05	0.03	0.05	0.06	1	0.73
	$\widehat{\theta}_{ML}$	0.03	0.05	0.02	0.04	1	0.60	0.03	0.05	0.02	0.07	1	0.74
	$\widehat{\theta}_{\delta_S}$	**0.02**	**− 0.1**	**5e−3**	**8e−3**	**0.93**	**0.07**	**0.02**	**− 0.10**	**0.01**	**0.02**	**0.91**	**0.10**
	$\widehat{\theta}$	**0.03**	**− 0.06**	**0.02**	**0.01**	**0.92**	**0.08**	**0.03**	**− 0.06**	**0.02**	**0.03**	**0.87**	**0.12**
$\Psi_{\delta_{Ab}}$	$\widehat{\theta}^{ML}_{\delta_S}$	0.11	0.18	0.08	0.02	1	0.10	0.33	0.41	0.17	0.05	1	0.56
	$\widehat{\theta}_{ML}$	0.20	0.29	0.11	0.02	1	0.50	0.30	0.34	0.19	0.05	1	0.69
	$\widehat{\theta}_{\delta_S}$	**0.10**	**− 0.30**	**0.01**	**0.01**	**0.95**	**0.03**	**0.10**	**− 0.16**	**0.08**	**0.06**	**0.91**	**0.04**
	$\widehat{\theta}$	**0.11**	**− 0.27**	**0.04**	**0.04**	**0.95**	**0.04**	**0.15**	**− 0.29**	**0.06**	**0.10**	**0.88**	**0.06**

estimate it. Regarding population parameters, we are particularly interested in the behavior of estimation methods for θ_{δ_S} and θ_{ϕ_S}. Indeed, a parameter sensitivity analysis shows the symmetric role of θ_{δ_S} and θ_{ϕ_S} on the ODE solution (see Balelli et al. (2020)). Thus, they are likely to face practical identifiability problems. To

investigate this effect, we estimate the parameters when $\theta^*_{\delta_S}$ 1. is known (the corresponding estimators will be denoted with the subscript δ_S), or 2. has to be estimated as well.

3.3.1 Well-specified case

We generate $n_i = 11$ longitudinal observations on the interval $[0, T] = [0, 364]$ with measurement noise of standard deviation $\sigma^* = 100$. For each subject i, the initial condition has been generated according to $A^*_{i,0} \sim N(\overline{A_0}, \sigma^2_{\overline{A_0}})$ with $\overline{A_0} = 500$ and $\sigma_{\overline{A_0}} = 260$ to reflect the dispersion observed in data presented in Pasin et al. (2019). We present the estimation results in Table 4—left side.

Our method gives an improved estimation with a dramatically reduced variance for $\theta^*_{\delta_S}$ comparing to ML, as well as an improved estimate for the $\left\{b^*_i\right\}_{i=1,\dots n}$ in all cases. We assume that is due to the committed estimation error for θ^* which causes model error during $\left\{b^*_i\right\}_{i=1,\dots n}$ estimation, not accounted for by ML. This in turn explains why variance Ψ^* is better estimated with our approach. In this mixed-effect context, this cause of model error is systematically present and claims for the use of estimation methods taking it into account when subject specific parameters are critical for the practitioner.

3.3.2 Misspecified case in presence of model error at the subject level

The data are generated with a stochastic perturbed version of ODE (17):

$$dA_i = \left(\phi_{S,i} e^{-\delta_S t} + \phi_{L,i} e^{-\delta_L t} - \delta_{Ab,i} A_i\right) dt + \alpha dB_t \tag{18}$$

where B_t is a Wiener process and $\alpha = 10$ its diffusion coefficient. The value for α has been chosen big enough to produce significantly perturbed trajectories but small enough to ensure that ODE (17) is still a relevant approximation for estimation purpose. The results are presented in Table 4—right side. Our method outperforms the ML for $\theta^*_{\delta_S}$ as well as for $\left\{b^*_i\right\}_{i\in[1,n]}$ estimation and their variances. However, we acknowledge that this last simulation setting is challenging even for our approach

Table 5 Estimation presented in Pasin et al. (2019) and via our approach

	Estimations from Pasin et al. (2019)	Optimal Control appraoch
Parameter	Mean IC95%	Mean IC95%
θ_{δ_S}	−0.57 [−1.02, −0.02]	−0.18 [−0.58, 0.22]
θ_{ϕ_S}	7.92 [7.52, 8.30]	7.45 [6.85, 7.96]
θ_{ϕ_L}	2.78 [2.62, 3.01]	2.58 [2.15, 3.01]
$\theta_{\delta_{Ab}}$	−3.54 [−3.62, −3.45]	−3.48 [−3.95, −3.01]
Ψ_{ϕ_S}	0.92 [0.83, 1.01]	0.64 [0.60, 0.70]
Ψ_{ϕ_L}	0.85 [0.78, 0.92]	0.70 [0.55, 0.90]
$\Psi_{\delta_{Ab}}$	0.30 [0.24, 0.36]	0.25 [0.19, 0.31]

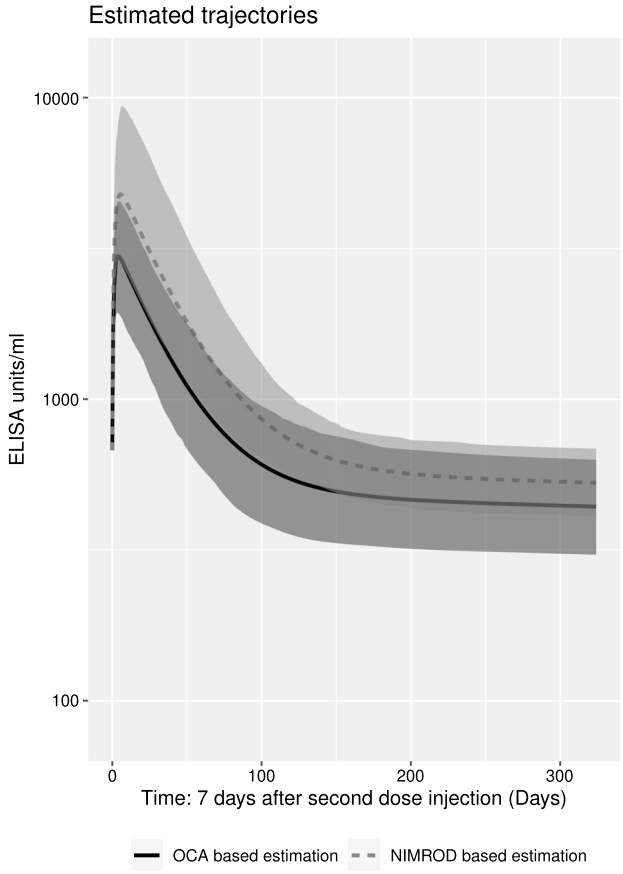


Fig. 2 Mean trajectory for Pasin et al. (2019) estimation (Dashed line) and the optimal control approach estimation (Solid line). Shaded area are the 95% confidence intervals

with confidence coverage around 90% for most of parameters, below the theoretical rate of 95%.

4 Real data analysis

We use the presented estimation approach to address the same problem as Pasin et al. (2019). This real data example is similar to the synthetic scenario performed in Sect. 3.3. In brief, we use data from a phase I trial in East Africa evaluating the effect of an heterologous anti-Ebola vaccine strategy in which Ad26.ZEBOV was injected first and then MVA-BN-Filo with a delay of 28 days between the two doses. We consider a population of n=28 individuals, with in average 5 measurements per subject. In order to ensure a fair comparison, we adopt a

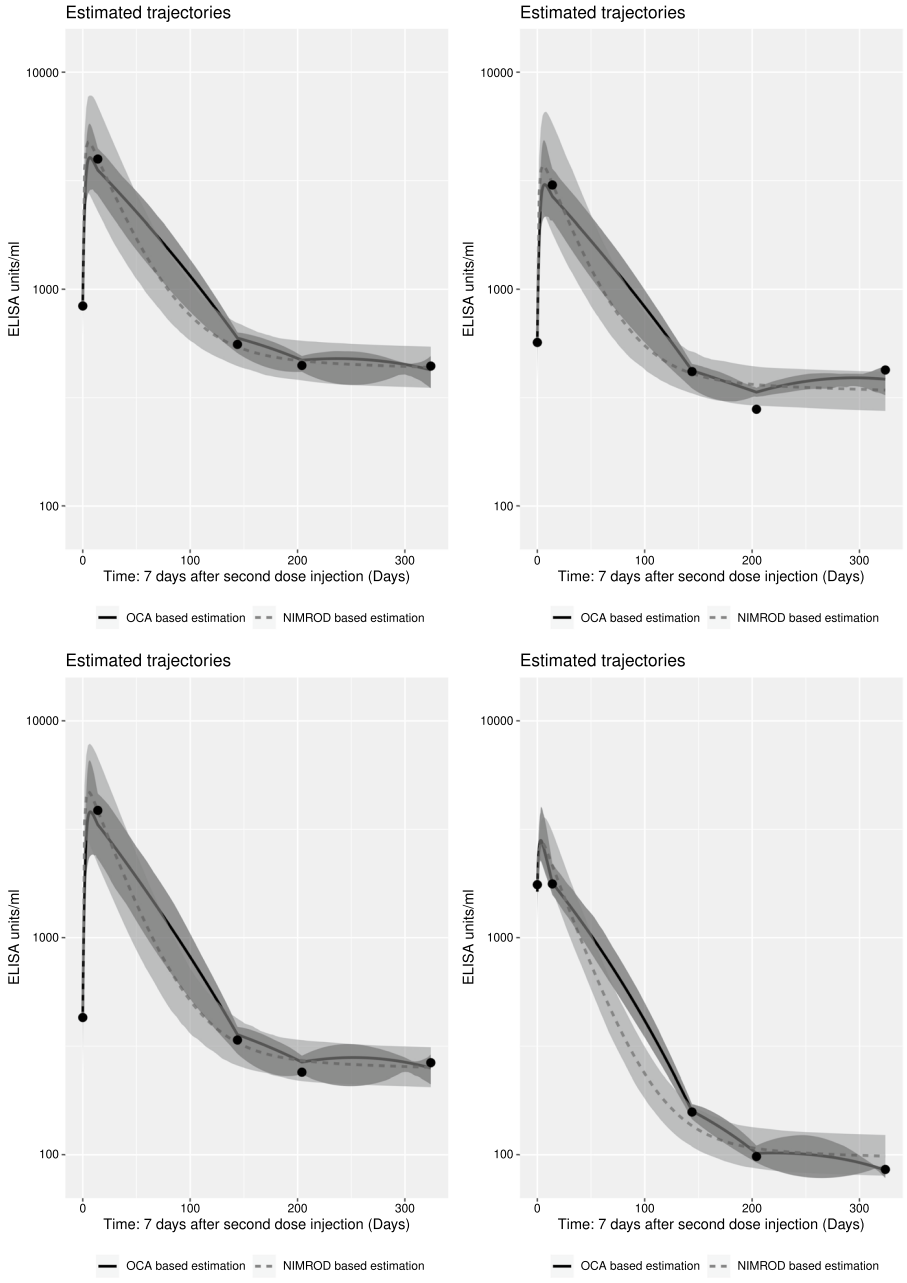


Fig. 3 Examples of fitted trajectories for both methods for four different random subjects. Dashed lines: fitted ODE solutions from Pasin et al. (2019). Solid line: optimal trajectories $\overline{X}_{\hat{\theta},\hat{b}_i}$ obtained with optimal control approach. Shaded area are the 95% confidence intervals

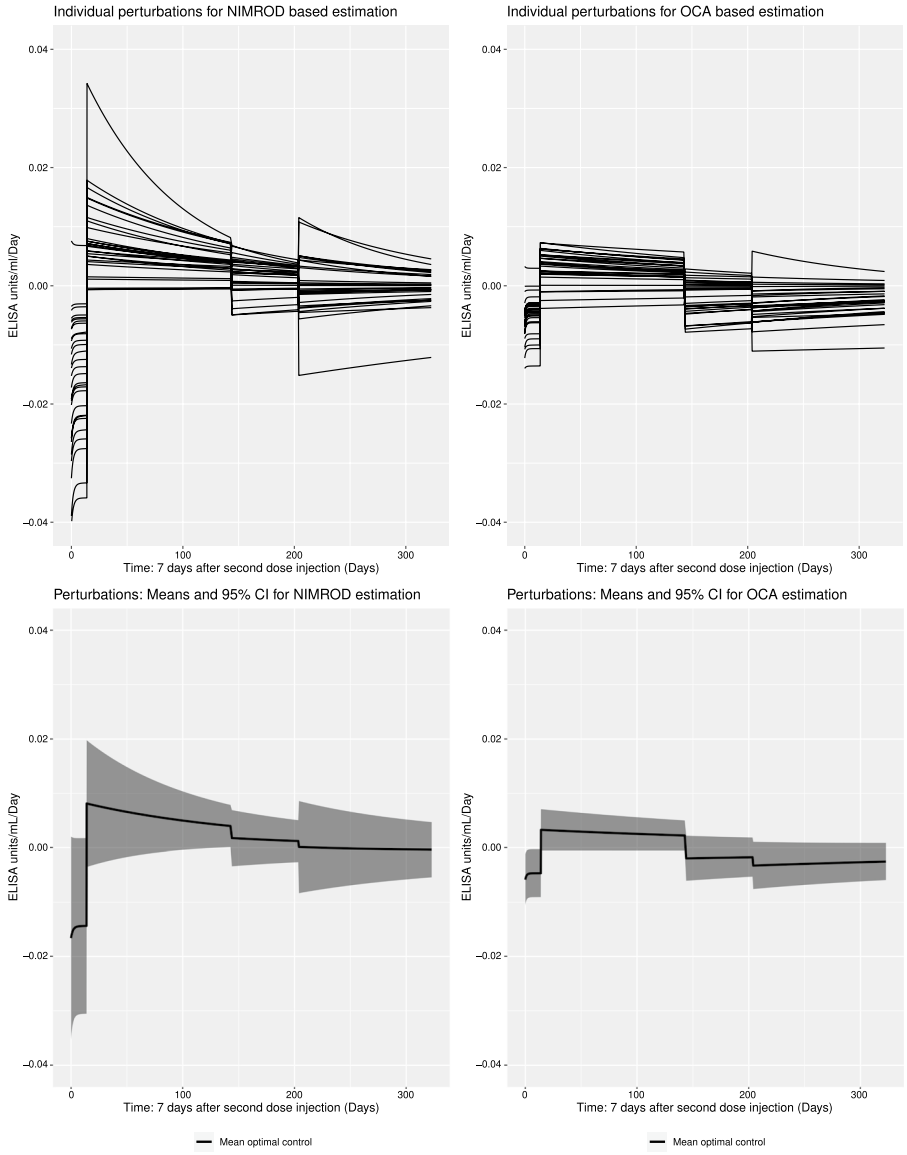


Fig. 4 (1) Up: Estimated residual controls for each subject, (2) bottom: mean optimal control and 95% confidence interval for the optimal controls a) left: $\overline{u}_{i,\hat{\theta}^P,\hat{b}_i^P,y_{i,0}}$ obtained from parameter estimation in Pasin et al. (2019), b) right: $\overline{u}_{i,\hat{\theta},b_i(\hat{\theta})}$ obtained from our estimation

Bayesian framework for $\theta = \left(\theta_{\delta_S}, \theta_{\phi_S}, \theta_{\phi_L}, \theta_{\delta_{Ab}}\right)$ and used the same prior distribution as in the original paper:

$$\pi(\theta) \sim N\left(\begin{pmatrix} -1 \\ 0 \\ 0 \\ -4.1 \end{pmatrix}, \begin{pmatrix} 25 & 0 & 0 & 0 \\ 0 & 100 & 0 & 0 \\ 0 & 0 & 100 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\right).$$

We set our mesh-size to get 200 discretization points for each subject and we use $U = 10$ i.e., a value lower than in the simulated data case because of the presence of model error. We also proceed to the log-transformation of the data to stabilize the measurement noise variance. Using the transformation $\widetilde{A}_i(t) := \log_{10} A_i(t)$ in Equation (17), this drives us to use the following nonlinear model:

$$\dot{\widetilde{A}}_i(t) = \frac{1}{\ln(10)}\left(\phi_{S,i}e^{-\delta_S t} + \phi_{L,i}e^{-\delta_L t}\right)10^{-\widetilde{A}_i(t)} - \frac{\delta_{Ab,i}}{\ln(10)}. \tag{19}$$

We choose $A_{\theta,b_i}(t,x) = \frac{1}{\ln(10)}\left(\phi_{S,i}e^{-\delta_S t} + \phi_{L,i}e^{-\delta_L t}\right)\frac{10^{-x}}{x}$ and $r_{\theta,b_i}(t) = -\frac{\delta_{Ab,i}}{\ln(10)}$ for the pseudo-linear formulation of the model. In Table 5, we compare our estimations with those presented in Pasin et al. (2019) obtained using the NIMROD software (Prague et al. 2013). Both methods produce estimations with overlapping confidence intervals for θ supporting the previous published results in term of antibodies concentrations dynamics over time. Still, significant differences appear for $\left(\Psi_{\phi_S}, \Psi_{\phi_L}, \Psi_{\delta_{Ab}}\right)$ estimation with lower dispersion of random effects in the optimal control approach. This is explained by the fact that a part of the variability is now carried out by subject-specific perturbations $\left\{\overline{u}_{i,\hat{\theta},b_i(\hat{\theta})}\right\}_{i=1,\dots n}$. In Fig. 2, we plot the mean curve for both estimation methods, that is the solution of ODE (19) with θ value given by Table 5 and random effects set to 0. The mean evolution are comparable between the two approaches. This is confirmed at the individual level in Fig. 3.

Finally, our method can be used to assess the model adequacy via the temporal evolution analysis of $\left\{\overline{u}_{i,\hat{\theta},b_i(\hat{\theta})}\right\}_{i=1,\dots n}$ estimated as byproducts of our method. In Sect. 2.1, we have also indicated that perturbations $\overline{u}_{i,\theta,b_i,x_{i,0}}$ can be computed for an arbitrary set $(\theta, b_i, x_{i,0})$. In particular, we estimate $\left\{\overline{u}_{i,\hat{\theta}^P,\hat{b}_i^P,y_{i,0}}\right\}_{i=1,\dots n}$, the committed error corresponding to $(\hat{\theta}^P, \hat{b}_i^P)$, the population and subject specific estimators obtained in Pasin et al. (2019). In Fig. 4, we plot both perturbation sets. Our method leads to residual perturbations of smaller magnitudes and narrower confidence intervals. This means that our approach produces an estimation which minimizes the committed model error for each subject comparing to a method based only on a data fitting criterion as in Pasin et al. (2019). Moreover, by reducing the size of the confidence interval for estimated perturbations, we conclude to a mean perturbation among the population which is statistically different from zero at the beginning of observation interval. This may indicate presence of model misspecification.

5 Conclusion

In this paper, we propose an estimation method that addresses problems encountered by classical approaches in NLME−ODE models. We identify the following shortcomings for exact methods such as likelihood-based inference: their difficulty in the presence of model misspecification, their need to estimate initial conditions as regular random effects, and their dramatic performance degradation in the presence of poorly identifiable parameters. We propose here a method based on an approximation of the profiled likelihood and control theory that accounts for the presence of potential model uncertainty at the subject level and that can be easily profiled on initial conditions. Simulations with and without model error illustrate the advantages of regularization techniques for estimating poorly identifiable parameters, subject-specific parameters, and their variances in NLME−ODEs. In addition, bypassing the estimation of initial conditions represents a clear advantage for partially observed systems comparing to likelihood based approaches, as emphasized in the simulations.

Still, this benefit in term of estimation accuracy comes with a computational price. On a server (see https://plafrim-users.gitlabpages.inria.fr/doc/ for more server details) with the parallelization package Snow in R language, and for a given choice of penalization matrix U, it takes approximately 10–15 min to obtain an estimation for the two-dimensional linear model, 30 min for the insulin model and 3-4 h for the antibody concentration evolution one, whereas it was a matter of minutes for the other approaches. Nevertheless, the use of compiled languages and proper parallelization could reduce the computation time. Moreover, we have willingly separated the formal definition of the optimal control problem required by our method and the numerical procedure used to solve it, in case it may exist better suited approaches for this specific control problem. Right now, our current strategy allows us to profile on initial conditions (despite requiring continuous observations and thus excluding applications to count or binary ones), therefore looking for another numerical procedure is beyond the scope of this paper.

Finally, the qualitative assessment of model misspecification exposed in Sect. 4 can be made more rigorous. In a one subject setting, the estimation of a perturbation term at the derivative level via non-parametric procedures to test model error presence has been already explored (Hooker et al. 2015; Engelhardt et al. 2017). Comparing to statistical methods solely based on data fitting criteria, they generally produce more sensitive statistical tests and can explore misspecification presence even for unobserved state−variables. Our control based approach can extend such tests to a population framework, while avoiding issues due to hyperparameter selection required for non-parametric statistical methods which can appear for a growing number of subjects. For example, to stay in a Bayesian setting, we can specify a prior distribution for the controls and then compare it with the obtained posterior once the inference is made. This would lead to a semi-parametric inference problem for which an optimal control based approach has already been proven useful (see Clairon (2020)). This is a subject for further work.

6 Supplementary information

A supplementary file containing the appendixes and proof of theorem 1 is available alongside this article.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00180-023-01420-x.

Acknowledgements Experiments presented in this paper were carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d' Aquitaine (see https://www.plafrim.fr/). This manuscript was developed under WP4 of EBOVAC3.

Funding This work has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under projects EBOVAC1 and EBOVAC3 (respectively grant agreement No 115854 and No 800176). The IMI2 Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Association.

Availability of data and materials Not applicable

Code availability Our estimation method is implemented in R and a code reproducing the examples of Sect. 3 is available on a GitHub repository located here (https://github.com/QuentinClairon/NLME_ODE_estimation_via_optimal_control.git).

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Aliyu M (2011) Nonlinear H-Infinity Control, Hamiltonian Systems and Hamilton-Jacobi Equations, CRC Press

Balelli I, Pasin C, Prague M et al (2020) A model for establishment, maintenance and reactivation of the immune response after vaccination against Ebola virus. J Theor Biol 495:110254

Brynjarsdottir J, O'Hagan A (2014) Learning about physical parameters: the importance of model discrepancy. Inverse Prob 30:24

Campbell D (2007) Bayesian collocation tempering and generalized profiling for estimation of parameters from differential equation models. PhD thesis, McGill University Montreal, Quebec

Cimen T (2008) State−dependent Riccati equation (SDRE) control: a survey. IFAC Proc 41:3761–3775

Cimen T, Banks S (2004) Global optimal feedback control for general nonlinear systems with nonquadratic performance criteria. Syst Control Lett 53:327–346

Clairon Q (2020) A regularization method for the parameter estimation problem in ordinary differential equations via discrete optimal control theory. J Stat Plan Inference 210:1–9

Clarke F (2013) Functional analysis, calculus of variations and optimal control, Graduate Texts in Mathematics, Springer, London

Comets E, Lavenu A, Lavielle M (2017) Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. J Stat Softw 80:1–42
De Gaetano A, Arino O (2000) Mathematical modelling of the intravenous glucose tolerance test. J Math Biol 40(2):136–168
Donnet S, Samson A (2006) Estimation of parameters in incomplete data models defined by dynamical systems. J Stat Plan Inference 137(9):2815–2831
Engelhardt B, Kschischo M, Fröhlich H (2017) A Bayesian approach to estimating hidden variables as well as missing and wrong molecular interactions in ordinary differential equation-based mathematical models. J R Soc Interface 14(131):20170332
Engl H, Flamm C, Kügler P et al (2009) Inverse problems in systems biology. Inverse Prob 25(12):123014
Hooker G, Ellner SP, Roditi LD, Earn DJ (2011) Parameterizing state−space models for infectious disease dynamics by generalized profiling: measles in Ontario. J R Soc 8:961–974
Guedj J, Thiebaut R, Commenges D (2007) Maximum likelihood estimation in dynamical models of HIV. Biometrics 63:1198–206
Gutenkunst RN, Waterfall J, Casey F et al (2007) Universally sloppy parameter sensitivities in systems biology models. Public Libr Sci Comput Biol 3:e189
Hooker G, Ellner SP et al (2015) Goodness of fit in nonlinear dynamics: Misspecified rates or misspecified states? Ann Appl Stat 9(2):754–776
Huang Y, Dagne G (2011) A Bayesian approach to joint mixed-effects models with a skew normal distribution and measurement errors in covariates. Biometrics 67:260–269
Huang Y, Lu T (2008) Modeling long-term longitudinal HIV dynamics with application to an aids clinical study. Ann Appl Stat 2:1348–1408
Kampen NV (1992) Stochastic process in physics and chemistry. Elsevier
Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. J R Stat Soc Ser B (Stat Methodol) 63(3):425–464
Kurtz T (1978) Strong approximation theorems for density dependent Markov chains. Stoch Process Appl 6:223–240
Lavielle M, Aarons L (2015) What do we mean by identifiability in mixed effects models? J Pharmacokinet Pharmacodyn 43:111–122
Lavielle M, Mentré F (2007) Estimation of population pharmacokinetic parameters of saquinavir in HIV patients with the monolix software. J Pharmacokinet Pharmacodyn 34:229–249
Leary TO, Sutton A, Marder E (2015) Computational models in the age of large datasets. Curr Opin Neurobiol 32:87–94
Lunn D, Thomas A, Best N et al (2000) Winbugs—a Bayesian modelling framework: concepts, structure and extensibility. Stat Comput 10:325–337
Lavielle M, Samson A, Karina Fermin A, Mentre F (2011) Maximum likelihood estimation of long terms HIV dynamic models and antiviral response. Biometrics 67:250–259
Murphy S, der Vaart AV (2000) On profile likelihood. J Am Stat Assoc 95:449–465
Nash JC (2016) Using and extending the optimr package
Pasin C, Balelli I, Van Effelterre T et al (2019) Dynamics of the humoral immune response to a prime−boost Ebola vaccine: quantification and sources of variation. J Virol 93(18):e00579-19
Perelson A, Neumann A, Markowitz M et al (1996) Hiv-1 dynamics in vivo: virion clearance rate, infected cell life−span, and viral generation time. Science 271:1582–1586
Pinheiro J, Bates DM (1994) Approximations to the loglikelihood function in the nonlinear mixed effects model. J Comput Graph Stat 4:12–35
Prague M, Commenges D, Drylewicz J et al (2012) Treatment monitoring of HIV-infected patients based on mechanistic models. Biometrics 68(3):902–911
Prague M, Commengues D, Guedj J et al (2013) Nimrod: a program for inference via a normal approximation of the posterior in models with random effects based on ordinary differential equations. Comput Methods Programs Biomed 111:447–458
Raftery A, Bao L (2010) Estimating and projecting trends in HIV/aids generalized epidemics using incremental mixture importance sampling. Biometrics 66:1162–1173
Ramsay J, Hooker G, Cao J et al (2007) Parameter estimation for differential equations: a generalized smoothing approach. J R Stat Soc 69:741–796
Sartori N (2003) Modified profile likelihood in models with stratum nuisance parameters. Biometrika 90:533–549
Sontag E (1998) Mathematical control theory: deterministic finite−dimensional systems. Springer, New York

Tornoe C, Agerso H, Jonsson EN et al (2004) Non-linear mixed-effects pharmacokinetic/pharmacodynamic modelling in NLME using differential equations. Comput Methods Programs Biomed 76:31–41

Tuo R, Wu C (2015) Efficient calibration for imperfect computer models. Ann Stat

van der Vaart A (1998) Asymptotic Statistics, Cambridge Series in Statistical and Probabilities Mathematics, Cambridge University Press

Varah JM (1982) A spline least squares method for numerical parameter estimation in differential equations. SIAM J Sci Stat Comput 3(1):28–46

Villain L, Commenges D, Pasin C et al (2019) Adaptive protocols based on predictions from a mechanistic model of the effect of IL7 on CD4 counts. Stat Med 38(2):221–235

Wang L, Cao J, Ramsay J et al (2014) Estimating mixed-effects differential equation models. Stat Comput 24:111–121

6.2 Prague et al. 2017 (Biometrics) Causal properties of Mechanistic Models

Dynamic versus marginal structural models for estimating the effect of HAART on CD4 in observational. Prague M. Commenges D., Gran JM., Ledergerber B., Young J., Furrer H. and Thiébaut R. *Biometrics.* 73(1) - 294-304 - July 2016.

This article integrates in the first axis of my research "Statistical Methodology for Estimations and Building of Mechanistic Models". It has been initiated during a short (2 months) research visit in University of Oslo between my PhD graduation and the start of my postdoctoral fellowship.

I selected this paper because it highlights the real strength of mechanistic models in comparison to more descriptive approaches. It offers a comprehensive catalog of methods for analyzing observational studies with informative treatment initiation, a challenging area due to the tendency to start treatment in more severe cases. The paper effectively reconciles and contrasts Marginal Structural Models with mechanistic models, underscoring that mechanistic models are a valid approach to causal inference. A primary challenge in this work was accessing data and establishing a simulation pipeline that didn't favor one approach over the other.

Biometrics 73, 294–304
March 2017

DOI: 10.1111/biom.12564

Dynamic Models for Estimating the Effect of HAART on CD4 in Observational Studies: Application to the Aquitaine Cohort and the Swiss HIV Cohort Study

Mélanie Prague,[1,*] Daniel Commenges,[2,3,4] Jon Michael Gran,[5] Bruno Ledergerber,[6] Jim Young,[7] Hansjakob Furrer,[8] and Rodolphe Thiébaut[2,3,4]

[1]Harvard T.H. Chan School of Public Health, Biostatistics Department, Boston, U.S.A.
[2]University of Bordeaux, ISPED, F-33000 Bordeaux, France
[3]INSERM, U1219 Bordeaux Population Health Research Centre, F-33000, Bordeaux, France
[4]INRIA (SISTM) Centre Recherche Bordeaux Sud-Ouest, University of Bordeaux, Talence, France
[5]Oslo Center for Biostatistics and Epidemiology, Oslo University Hospital and University of Oslo, Norway
[6]Division of Infectious Diseases and Hospital Epidemiology, University Hospital of Zurich, Switzerland
[7]Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital of Basel, Switzerland
[8]Department of Infectious Diseases Bern University Hospital, University of Bern, Switzerland
[*]*email:* mprague@hsph.harvard.edu

Summary. Highly active antiretroviral therapy (HAART) has proved efficient in increasing CD4 counts in many randomized clinical trials. Because randomized trials have some limitations (e.g., short duration, highly selected subjects), it is interesting to assess the effect of treatments using observational studies. This is challenging because treatment is started preferentially in subjects with severe conditions. This general problem had been treated using Marginal Structural Models (MSM) relying on the counterfactual formulation. Another approach to causality is based on dynamical models. We present three discrete-time dynamic models based on linear increments models (LIM): the first one based on one difference equation for CD4 counts, the second with an equilibrium point, and the third based on a system of two difference equations, which allows jointly modeling CD4 counts and viral load. We also consider continuous-time models based on ordinary differential equations with non-linear mixed effects (ODE-NLME). These mechanistic models allow incorporating biological knowledge when available, which leads to increased statistical evidence for detecting treatment effect. Because inference in ODE-NLME is numerically challenging and requires specific methods and softwares, LIM are a valuable intermediary option in terms of consistency, precision, and complexity. We compare the different approaches in simulation and in illustration on the ANRS CO3 Aquitaine Cohort and the Swiss HIV Cohort Study.

Key words: Dynamic mechanistic models; Linear increment models (LIM); Marginal structural models (MSM); Non-linear mixed effect models (NLME); Observational study; Ordinary differential equation (ODE).

1. Introduction

Randomized clinical trials often have short durations and include highly selected subjects, thus assessing the effect of a treatment using observational studies is useful. However, it is challenging because the treatment may change, and covariate history of a subject up to time t may influence treatment given after t, and may also influence the outcome of interest, which induces a time-dependent confounding. For instance, one may wish to assess the effect of antiretroviral therapy in HIV infected subjects. As CD4+ T-lymphocytes (CD4, in short) are the main target cells of the HIV virus, it is possible to assess the effect of a treatment on the blood concentration of these cells: CD4 counts are measurements of this concentration. In observational studies, however, the decision to start an antiretroviral therapy may depend on CD4 counts as well as on other covariates. In this setting, it has been demonstrated that a conventional regression analysis leads to biased estimates of the treatment effect, typically underestimating it, and possibly (wrongly) indicating a negative effect. This is called "confounding by indication" (Walker, 1996).

Marginal structural models (MSM) (Robins et al., 2000) have been proposed for dealing with this issue; this is based on choosing a causal model in terms of potential responses, which are often counterfactual, to the different treatment histories. The parameters of a MSM can be estimated through a weighted approach but other methods exist (Petersen et al., 2006). The weights are the inverse probability of treatment assignment and are obtained through a "treatment model" which includes the covariates linked to the outcome. Because data are correlated, we use an inverse probability weighted generalized estimating equation (GEE). This approach has been applied by Hernán et al. (2002) and Cole et al. (2005) for estimating the effect of zidovudine and of highly active antiretroviral therapy (HAART) on CD4 count. Sterne et al. (2005) and Cole et al. (2007) used it for estimating the effect of HAART on viral load and on AIDS or death.

An alternative view to causality that does not use the potential responses representation is to use dynamic models. This has been pioneered by Granger (1969), Dawid (2000), and further developed by Didelez (2008), Commenges and Gégout-Petit (2009), Gégout-Petit and Commenges (2010), and Eichler and Didelez (2010). Assumptions needed for a causal interpretation of dynamic models have been presented in Arjas and Parner (2004) and Commenges and Gégout-Petit (2015). Dynamical models in discrete time, and in particular linear increment models (LIM), have been proposed by Diggle et al. (2007) and Hoff et al. (2014). Aalen et al. (2012) have suggested that such models can be useful for studying the HAART effect on CD4 counts or viral load. Discrete-time models, however, may not be completely satisfactory because the processes of interest most often exist in continuous time. Systems of differential equations in continuous time can also be used to model the interaction between HIV and CD4 cells populations. Models based on differential equations, called "mechanistic," considerably helped in understanding some important features of the infection: see Perelson (2002) for a review. In our setting, it is possible to model the treatment effect from a biological perspective. Introducing random effects is an efficient way to model differences between subjects with a minimum of additional parameters (Wu, 2005; Guedj et al., 2007; Lavielle et al., 2011). Mechanistic models have mostly been used to analyze data from clinical trials. Using mechanistic models to estimate the effect of HAART based on data of large observational cohorts is possible, but to our knowledge, has never been attempted.

The aim of this article is to propose dynamic models in discrete and continuous time for assessing the causal effect of a treatment on a marker in observational studies. Specifically, we aim to estimate the HAART effect in HIV infected patients. We present several possible dynamic models, as well as MSMs, and compare them using simulations and real data. Although we worked out analytics in simple cases (see Appendix), the comparison is mainly empirical and aims to describe the assumptions and variability of results provided by the various methods. In Section 2, we present the statistical models: the naive model, the MSMs, the discrete-time dynamic models, and the mechanistic approach. In Section 3, we compare the results of these models in simulation, where the data are generated from a complex mechanistic model. Section 4 is the application on the data of two cohorts of HIV infected patients: the Swiss HIV Cohort Study (SHCS) and the ANRS CO3 Aquitaine cohort. Section 5 concludes.

2. Modeling the Treatment Effect in Observational Studies

2.1. *Notations and the Naive Model*

We denote the value of a physiological marker for subject i at time t by Y_t^i, say the CD4 count. The value of a treatment given at time t to subject i is denoted by A_t^i, say HAART. For sake of simplicity, we only model two treatment states: $A_t^i = 0$ when treatment is not given, and $A_t^i = 1$ when treatment is given, and we assume that once initiated the treatment is not interrupted; however, one could expand the model to accommodate different levels of treatment. If treatment is started at time t then $A_t^i = 1$ and $A_{t-1}^i = 0$. We use overbars to represent the histories of these processes: for instance $\overline{\boldsymbol{A}_t^i} = (A_0^i, A_1^i, \ldots, A_t^i)$. We denote by $\text{cum}(\overline{\boldsymbol{A}_t^i}) = \sum_{k=1}^{t} A_k^i$ the cumulative time under treatment until time t. In the absence of confounding by indication, the simplest model would be to regress Y_t^i on $\text{cum}(\overline{\boldsymbol{A}_{t-1}^i})$. Cole et al. (2005) notes the advantages of a piecewise linear regression model that allows a change in the effect of treatment after 1 year. Thus, our Model 1 is a naive model with different treatment effects before and after 1 year:

$$\text{E}(Y_t^i|\overline{\boldsymbol{A}_{t-1}^i}) = \beta_0 + \beta_1 \text{cum}(\overline{\boldsymbol{A}_{t-1}^i}) + \beta_2 \text{cumlag}(\overline{\boldsymbol{A}_{t-1}^i}). \quad (1)$$

where $\text{cumlag}(\overline{\boldsymbol{A}_{t-1}^i})$ is the cumulative time under treatment up to time t minus 1 year: $\text{cumlag}(\overline{\boldsymbol{A}_{t-1}^i}) = \max(0, \text{cum}(\overline{\boldsymbol{A}_{t-1}^i}) - 1)$ (with the convention $A_t^i = 0$ for $t < 0$). The β's can be estimated by conventional GEE (Liang and Zeger, 1986) because we are interested in the population average. We use the independence working correlation structure in our analyses; otherwise results could be biased because of the presence of time-varying covariates (Pepe and Anderson, 1994).

2.2. *Marginal Structural Models (MSM)*

Because treatment is given to subjects with low CD4 counts, treated subjects tend to have low CD4 counts. Thus, the true value of parameter β_1 in Model 1 cannot be interpreted as the causal effect. MSMs are designed to estimate the causal effect of a treatment given time dependent confounding. It is assumed that to each particular value of treatment history $\overline{\boldsymbol{a}_t^i}$ of $\overline{\boldsymbol{A}_t^i}$, a potential outcome $Y_t^i(\overline{\boldsymbol{a}_t^i})$ is associated (possibly contrary to the treatment actually received). A model is postulated to describe how the potential outcomes vary as a function of the different treatment trajectories. Hernán et al. (2002) and Cole et al. (2005) proposed benchmark models that also adjust for confounders such as time (t) and baseline value of biomarkers (Y_0) in the regression. To duplicate this, our Model 2 is:

$$\begin{aligned}\text{E}(Y_t^i(\overline{\boldsymbol{a}_t^i})|\overline{\boldsymbol{a}_{t-1}^i}, Y_0^i) = \beta_0 + \beta_1 \text{cum}(\overline{\boldsymbol{a}_{t-1}^i}) + \beta_2 \text{cumlag}(\overline{\boldsymbol{a}_{t-1}^i}) \\ + \beta_3 t + \beta_4 Y_0^i. \quad (2)\end{aligned}$$

Robins et al. (2000) showed that the causal parameters of this model can be estimated with a suitably weighted GEE. The weights represent the inverse probability of treatment. The probability of treatment at time t depends on the history up to time t of a vector of variables $\boldsymbol{L}$ denoted $\overline{\boldsymbol{L}_t^i} = (\boldsymbol{L}_0^i, ..., \boldsymbol{L}_t^i)$; $\boldsymbol{L}_t^i$ typically includes include Y_t^i. The probability of treatment is estimated at each point in time using a treatment model (generally a logistic model) and the weights are the product over time of these probabilities; one often use stabilized weights as in equation (3). An extension allows for censoring (Cole et al., 2005; Cole and Hernán, 2008), however, the most important correction is generally for the probability of treatment (Ko et al., 2003). Stabilized inverse probability

of treatment weights are defined as:

$$SW^i(t) = \prod_{k=1}^{t} \frac{Pr(A_k^i = 1|\overline{\boldsymbol{A}_{k-1}^i}, \boldsymbol{L}_0^i)}{Pr(A_k^i = 1|\overline{\boldsymbol{A}_{k-1}^i}, \overline{\boldsymbol{L}_k^i})}. \quad (3)$$

Results from Model 2 are consistent if the treatment model corresponds to the true treatment assignment mechanism; i.e., if all the confounders (factors influencing both the outcome of interest and treatment assignment) have been taken into account. We defined $\boldsymbol{L}$ for treatment model in Model 2 as baseline and time-varying CD4 count in categories (<200, [200; 400], >400), viral load in categories (<401, 401 − 10,000, and >10,000), and an indicator of undetectable viral load. In Models 1 and 2 the effect of treatment on CD4 counts during the first year is given by β_1 and the effect after 1 year of treatment is given by $\beta_1 + \beta_2$.

2.3. *Modeling the Increment with a Dynamical MSM*

It is more relevant to model the change in the marker of interest, rather than its current value. This fits well with a causal thinking which considers that the change of a process depends on its present, and possibly past, states. Thus, we could model $Y_t^i - Y_{t-1}^i$; in order to account for non-equally spaced measurements, we might assume that the change in biomarker values is proportional to the time elapsed between two measurements; that is, $\Delta_t^i = c_t^i - c_{t-1}^i$, where c_t^i is the tth calendar time of observation since the baseline measure of subject i. We define the biomarkers increments as $Z_t^i = \frac{Y_t^i - Y_{t-1}^i}{\Delta_t^i}$. We take this approach when developing linear increment models in Section 2.4. However it is possible to express a MSM for the Z_t^i's as:

$$\mathrm{E}(Z_t^i(\overline{\boldsymbol{a}_t^i})|\overline{\boldsymbol{a}_{t-1}^i}, Y_0^i) = \beta_0 + \beta_1 a_{t-1}^i + \beta_2 a_{t-2}^i. \quad (4)$$

Our Model 3 is then the combination of this MSM for Z_t^i together with the treatment model (3). Note that with equally spaced measurements, Models 2 and 3 will be structurally identical if the same unit of time (e.g., year) is used for Δ_t^i and the cumulative exposures (see Appendix). Inference approaches will be different, thus, Model 3 should be more suitable than Model 2 because we use an independence working correlation matrix for the Z_t^i's, which is much more acceptable than assuming independence for the Y_t^i's.

2.4. *Full Discrete-Time Dynamical Models–Linear Increment Models (LIM)*

We can also fit a linear mixed-effects model for the Z_t^i's. The inter-subject variability is accounted for by a random effect $\boldsymbol{b}$ assumed normally distributed with zero expectation. Thus, LIM specify the distribution of $\boldsymbol{Z}^i$ conditional on the b_i. We propose three LIM; the simplest is Model 4, where the $\boldsymbol{\varepsilon}$'s are i.i.d. normal variables with zero expectation:

$$Z_t^i = \beta_0 + \beta_1 A_{t-1}^i + \beta_2 A_{t-2}^i + b_i + \varepsilon_t^i. \quad (5)$$

In both Models 3 and 4, the effect of treatment on the CD4 counts is $\beta_1 \bar{\Delta}_t$ during the first year of treatment, and $(\beta_1 + \beta_2)\bar{\Delta}_t$ after 1 year, where $\bar{\Delta}_t$ is the mean of all the Δ_t^i.

Many deterministic dynamical models have equilibrium points; similarly, many stochastic dynamical models tend toward a stationary process. This property fits very well with the behavior of biological systems since concentrations of many molecules or cells have a tendency to return to the same value, a property called "homeostasis." Difference equations of the type $Y_t^i - Y_{t-1}^i = \gamma_0 + \gamma_1 Y_{t-1}^i + \varepsilon_t^i$ correspond to an autoregressive model of order one, denoted AR(1): $Y_t^i = \gamma_0 + \gamma' Y_{t-1}^i + \varepsilon_t^i$ with $\gamma' = (\gamma_1 + 1)$. It is well known that if $|\gamma'| < 1$ this process converges toward a stationary process (in discrete time) with expectation $\mathrm{E}(Y_t^i) = -\frac{\gamma_0}{1-\gamma'} = -\frac{\gamma_0}{\gamma_1}$; this is always defined unless $\gamma_1 = 0$, as is the case in Model 4 which does not have a finite stationary expectation. When using a model which has this convergence property, it may not be necessary to have a two-slope model. For example, we define Model 5, which tends to a stationary process with expectation $-\frac{\beta_0+\beta_1}{\beta_2}$ for treated patients if $-2 < \beta_2 < 0$ and $\beta_0 + \beta_1 > 0$:

$$Z_t^i = \beta_0 + \beta_1 A_{t-1}^i + \beta_2 Y_{t-1}^i + b_i + \varepsilon_t^i. \quad (6)$$

A more realistic modeling of CD4 counts takes viral load into account. Here, we make a step toward mechanistic models because we know that the virus concentration and the CD4 concentration are inter-related processes. Thus, Model 6 is based on a system of two difference equations:

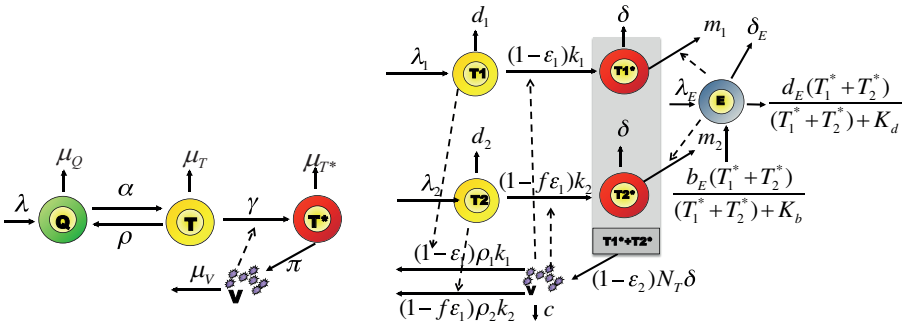
$$\begin{cases} Z_t^i = \beta_0 + \beta_1 A_{t-1}^i + \beta_2 Y_{t-1}^i + \beta_3 \mathrm{VL}_{t-1}^i + b_i + \varepsilon_{it}^1, \\ W_t^i = \alpha_0 + \alpha_1 A_{t-1}^i + \alpha_2 Y_{t-1}^i + \alpha_3 \mathrm{VL}_{t-1}^i + d_i + \varepsilon_{it}^2. \end{cases} \quad (7)$$

where $W_t^i = \frac{\mathrm{VL}_t^i - \mathrm{VL}_{t-1}^i}{\Delta_t^i}$, with VL_t^i the viral load at time t for patient i. The random effects d_i and b_i, and the errors ε_{it}^1 and ε_{it}^2 are all i.i.d. and normally distributed with zero expectation. In Models 5 and 6, one year and subsequent years increase in CD4, as well as the long-term change, are easily computed by solving the difference equations numerically. Moreover, for testing whether the treatment has an effect, it is convenient to test the hypotheses $\beta_1 = 0$ and $\alpha_1 = 0$, by Wald tests for instance. As a reviewer notes, an interesting question is the correspondence between LIM and MSM. The Appendix shows that under some (rather strong) assumptions, Model 4 estimates the same causal parameters as Models 2 and 3 without using a treatment model.

2.5. *Continuous Dynamical Models, Mechanistic Models (ODE-NLME)*

In reality, biomarkers processes exist in continuous time. A natural extension of a dynamic model in discrete time, as $\Delta_t^i \to 0$, is a model based on differential equations. (Perelson, 2002) give a review of some models for HIV dynamics based on ODE systems. In this article, we consider the "target cells model", called Model 7 and described bellow, which proved to provide a good fit and prediction abilities (Prague et al., 2013).

2.5.1. Biological system. We know that only infected cells (T^*) can produce viruses (V). The target cells model distinguishes between uninfected quiescent cells (Q) and target cells (T). The instantaneous change of concentrations of these populations at time t, for all real value of $t > 0$, is given by the



(a) Target cell model (b) Adams et al. (2005) model.

Figure 1. Mechanistic models for HIV dynamics. Type of cells of interest are viruses (V), effector cells E and CD4 cells which may be quiescent (Q), target cells (T, T_1, T_2) or infected (T^*, T_1^*, T_2^*). Parameters are defined in Table 1.

ODE system:

$$\begin{cases} \dfrac{\mathrm{d}Q_t^i}{\mathrm{d}t} = \lambda^i + \rho^i T_t^i - \alpha^i Q_t^i - \mu_Q^i Q_t^i, \\ \dfrac{\mathrm{d}T_t^i}{\mathrm{d}t} = \alpha^i Q_t^i - \gamma^i T_t^i V_t^i - \rho^i T_t^i - \mu_T^i T_t^i, \\ \dfrac{\mathrm{d}T^{*i}_t}{\mathrm{d}t} = \gamma^i T_t^i V_t^i - \mu_{T^*}^i T^{*i}_t, \\ \dfrac{\mathrm{d}V_t^i}{\mathrm{d}t} = \pi^i T^{*i}_t - \mu_V^i V_t^i. \end{cases} \tag{8}$$

The system is graphically represented in Figure 1a. Here, the parameters have biological meanings: λ is the production rate of new CD4 cells, the μ's are death rates of different populations of cells, α and ρ are transition rates between quiescent and target cells, π is the rate of production of virions by infected cells, and γ is the infectivity parameter. The model assumes that the rate of infection of target cells is γV_t.

2.5.2. Inter-individual variability. A model that allows parameters to vary between individual is a mixed effects model. We model the inter-individual variability on the log-transformed parameters denoted with a tilde to ensure positiveness during estimation. In this application, two random effects u_λ^i and $u_{\mu_{T^*}}^i$ are introduced (Prague et al., 2012): $\tilde{\lambda}^i = \tilde{\lambda}_0 + u_\lambda^i$ and $\tilde{\mu}_{T^*}^i = \tilde{\mu}_{T^*0} + u_{\mu_{T^*}}^i$. Biologically, the causal effect of treatment can be modeled as an effect on the infectivity parameter γ. The parameter γ depends on t through A_t^i, where we expect $\beta < 0$, so that the treatment decreases the infectivity of the virus:

$$\tilde{\gamma}^i(t) = \tilde{\gamma}_0 + \beta A_t^i. \tag{9}$$

2.5.3. Observation model. One important consequence of using continuous time models is that we must distinguish between the biological system which exists in continuous time and observations which are made at discrete times. To make an additive model for measurement error acceptable, we use 4th-root transformation for CD4 and a $\log_{10}$ transformation for the viral load. With errors ε_{ij}^1 and ε_{ij}^2 both i.i.d. and normally distributed, the observation model is:

$$\begin{aligned} (Y_j^i)^{1/4} &= [Q_{t_{ij}}^i + T_{t_{ij}}^i + T^{*i}_{t_{ij}}]^{1/4} + \varepsilon_{ij}^1 \ ; \\ \log_{10} \mathrm{VL}_{ij} &= \log_{10} V_{t_{ij}}^i + \varepsilon_{ij}^2. \end{aligned} \tag{10}$$

2.5.4. Inference. Inference is much more complex and computationally demanding than in discrete-time models. We base inference on a penalized maximum likelihood approach and to achieve identifiability, we include prior knowledge of mechanistic parameters. Our priors (Table 1) are based on estimates in the literature of these parameters (Prague et al., 2012). This approach has been implemented in the NIMROD program (Prague et al., 2013). Assessing the long-term treatment effect in Model 7 is possible by analytically computing the equilibrium point. One year and subsequent year increase in CD4 after treatment initiation can be computed by solving the ODE system for given values of the random effects. The marginal effect can be computed as the mean of the individual effects in the population. The β parameter in the infectivity definition gives the effect of treatment, and a Wald test can be used to test the no-effect hypothesis ($\beta = 0$).

3. Simulation Study

We simulated data with the Adams et al. (2005) model made up of two populations of target cells and a population of immune effectors such as cytotoxic T-lymphocytes (see Figure 1b and Table 1), which is much more complex than Model 7 (see Web-Supplementary Material A4). Individual variability was introduced by drawing parameters from normal distributions (with mean values listed in Table 1 and variances chosen to obtain a variation coefficient of 50%). By controlling the value of random effects, we ensured that the steady state baseline distributions of CD4 counts and viral load were consistent with the baseline values distributions found in Aquitaine cohort and SHCS data set. See Web-supplementary Material A1 for details. We generated observations every 3 months; the standard deviations of the measurement errors

Table 1

Meaning of parameters in the dynamical models presented in Figure 1. The upper part gives prior means and standard deviations for normal a priori distributions used for estimation of mechanistic parameters in Model 7 for the "Target cell model". The lower part gives parameter values used for data simulation from the Adams et al. (2005) model.

Name	Description		Normal priors used for analysis[a] on the log value of the parameter	
			Mean	SD.
λ	Natural production rate	$\frac{cells}{\mu L.day}$	2.55	1.90
μ_{T^*}	Natural death rate of T^* cells	$\frac{1}{day}$	−0.05	0.68
μ_Q	Natural death rate of Q cells	$\frac{1}{day}$	−9.00	1.00
α	Transition rates between Q and T cells	$\frac{1}{day}$	−4.00	2.00
ρ	Transition rates between T and Q cells	$\frac{1}{day}$	−4.34	1.38
μ_T	Virions natural death rate	$\frac{1}{day}$	−2.59	0.34
γ	Infectivity parameter	$\frac{\mu L}{day}$	−5.76	4.02
π	Rate of production of virions by infected cells	$\frac{1}{day}$	4.04	2.66
μ_V	Natural death rate of viruses	$\frac{1}{day}$	2.83	0.68

Name	Description	Units	Parameter Value used for simulations for each population (X)[b]			
			Type 1	Type 2	Effectors	Virus
λ_X	Natural production rate	$\frac{cells}{mL.day}$	5000	31.98	1.0	–
(1-ϵ_X)	Treatment efficacy	no unit	50%	83%	–	–
d_X	Natural death rate	$\frac{1}{day}$	0.01	0.01	0.25	–
δ_X	Infection-induced death rate	$\frac{1}{day}$	0.7	0.7	0.1	–
ρ_X	Number of virions infecting a cell	$\frac{virions}{cells}$	1	1	–	–
m_X	Immune-induced clearance rate	$\frac{mL}{cells.day}$	1×10^{-5}	1×10^{-5}	–	–
k_X	Infection rate	$\frac{mL}{virions.day}$	8×10^{-7}	1×10^{-4}	–	–
c	Virions natural death rate	$\frac{1}{day}$	–	–	–	13
N_T	Virions production per infected cells	$\frac{virions}{cells}$	–	–	–	100
K_b	Saturation constant cells birth	$\frac{cells}{mL}$	–	–	100	–
K_d	Saturation constant cells death	$\frac{cells}{mL}$	–	–	500	–
b_E	Infection-induced birth rate for E cells	$\frac{1}{day}$	–	–	0.3	–

[a]Reference and explanation for these choices can be found in Prague et al. (2012). [b]For each simulated patient, every parameter got a random effect leading to 50% coefficient of variation

Table 2

Estimated treatment effect on CD4 counts from simulated data: Model 1: Naive regression; Model 2: MSM on Y_t^i; Model 3: MSM on Z_t^i; Model 4: simple LIM; Model 5: autoregressive LIM; Model 6: LIM system; Model 7: mechanistic model.

	Simulated Data set with Adams et al. (2005) model						
		$n = 200$			$n = 1500$		
Model	β treatment[a]	Effect	Sd.	Z-stat[b]	Effect	Sd.	Z-stat[b]
Model 1	<1 yr	136	29	4.68	172	11	16.34
	>1 yr	−11	6	−1.99	−12	2	−5.07
	∞	$-\infty$	−	−	$-\infty$	−	−
Model 2	<1 yr	327	31	10.64	325	11	28.81
	>1 yr	−14	10	−1.45	−7	4	−1.56
	∞	$-\infty$	−	−	$-\infty$	−	−
Model 3	<1 yr	371	24	15.47	364	10	35.49
	>1 yr	8	3	2.45	4	1	3.2
	∞	$+\infty$	−	−	$+\infty$	−	−
Model 4	<1 yr	362	17	21.60	378	6	61.35
	>1 yr	8	12	0.7	7	4	1.61
	∞	$+\infty$	−	−	$+\infty$	−	−
Model 5	<1 yr	133	*	*	136	*	*
	>1 yr	84	*	*	86	*	*
	∞	359	*	*	370	*	*
	β_1	149	5	31.24	154	2	89.36
Model 6	<1 yr	325	*	*	334	*	*
	>1 yr	31	*	*	034	*	*
	∞ CD4	360	*	*	371	*	*
	∞ VL	−1.9	*	*	−2	*	*
	β_1	600	21	28.42	630	8	82.22
	α_1	−6.6	0.16	−40.86	−6.9	0.06	−120.51
Model 7	<1 yr	312	*	*	304	*	*
	>1 yr	2	*	*	4	*	*
	∞ CD4	308	*	*	306	*	*
	∞ VL	−5.6	*	*	−4.98	*	*
	β	−1.12	0.014	−79.3	−1.030	0.003	−295.6

[a]To be compared with mean treatment effect in treated for (<1 year; >1 year; ∞): benchmarks values are (350;12;370) for these simulations.

[b]Estimates for treatment effect (β) are significant at level 10% if the absolute value of Z-stat is greater than 1.64 and significant at level 5% if the absolute value of Z-stat is greater than 1.96.

*Simulated delta-method can lead to estimation of these values, but is not implemented here. Indeed, for Models 5, 6, and 7 significance of the treatment effect has to be evaluated through the mechanistic parameters β_1, α_1 and β.

were $\sigma_{VL} = 0.6$ and $\sigma_{CD4} = 0.1$. Viral load was artificially made undetectable at the level of 50 copies/mL. Treatment assignment was done by simulating a CD4 count assessment at every observation and by fixing a probability of treatment assignment depending on the observed CD4 count. We took empirical probabilities from the Aquitaine cohort and SHCS data set: treatment was attributed in 2, 28, or 47% of patients where their CD4 count was >400, [400, 200] or <200. No other confounder was considered. Simulated patients are supposed fully observed for 5 years. We simulated $n = 200$ and $n = 1500$ patients. Table 3 gives a general description of both simulated and real data sets and there is no obvious difference between simulated and real cohort data in their descriptive statistics. We define the "average causal effect in treated patients" as the mean difference between the observed CD4 according to the observed treatment initiation and the counterfactual CD4 under no treatment initiation. The result of this computation was a 350 cells increase in CD4 after 1 year, a 362 cells increase after 2 years and an overall increase of 370 CD4 cells after an infinite (large) time. Technical details and code for analysis are described in Web-Supplementary Material C.

Table 2 presents the estimates for Models 1–7 on the simulated data sets. The naive Model 1 largely underestimated the treatment effect. This was corrected by the MSM Models 2 and 3 (see details about weights in Web-Supplementary Material A2). Model 4 also yielded good estimates of the mean causal effect in treated patients. Moreover, it led to an increased significance of results compared to Model 3 for increase in CD4 count during the first year. Regarding increase in CD4 count in subsequent years, whereas the Model 3 underestimate its variability (for large samples, $12 \notin [2; 6]$), Model 4 is more reliable (for large samples, $12 \in [-1; 15]$). This underestimation of the variance of Model 3 may be driven by an overfitting of inverse probability of treatment weights. This overfitting could have arisen because the data

Table 3
Data description for illustrations : average viral load, CD4 counts, and percentage of treatment assignment in the population are displayed for simulated data and real data from the Aquitaine cohort and the SHCS. Statistics displayed are mean [Q1;Q3].

	Simulated data set		Aquitaine cohort	SHCS
n	200	1500	1591	1726
Missing data				
Administrative	–	–	81.6%	74.7%
Death	–	–	12.7%	6.4%
Lost of follow-up	–	–	5.7%	18.9%
CD4 count				
Baseline	428 [266 ; 545]	420 [253 ; 530]	471 [298 ; 612]	536 [357 ; 670]
Follow-up untreated	594 [485 ; 675]	588 [478 ; 656]	625 [440 ; 762]	543 [363 ; 675]
Follow-up treated	627 [417 ; 837]	606 [405 ; 801]	492 [315 ; 638]	507 [300 ; 660]
Viral Load				
Baseline	3.9 [3.3 ; 4.6]	4 [3.4 ; 4.7]	4.2 [3.6 ; 4.8]	4.0 [3.4 ; 4.6]
Follow-up untreated	3.5 [2.9 ; 4.2]	3.7 [3.1 ; 4.4]	3.3 [2.3 ; 4.2]	3.8 [3.1 ; 4.5]
Follow-up treated	2.6 [1.7 ; 3.2]	2.6 [1.7 ; 3.2]	2.7 [1.7 ; 3.6]	3.2 [2.4 ; 4.1]
% undetectable viral load (baseline,untreated, treated)	(3%;4%;40%)	(2%;3%;38%)	(7%;22%;48%)	(10%;15%;57%)
Treatment assignment				
Time (day)	412 [1 ; 631]	377 [91 ; 451]	727 [1 ; 1281]	548 [183 ; 752]
% treated	69%	65%	64%	34%

generation was based on CD4 cell count evolution and not on viral load, whereas the inverse probability of treatment weights were calculated from treatment models with both past CD4 cell count and viral load as predictors. We note that long-term increase in CD4 is infinite in Models 1–4, even though it happens at a really slow rate. On the contrary, Models 5–7 exhibit an equilibrium point which makes it possible to consider the long-term causal effect of treatment. All dynamic models gave an estimate of the long-term effect of treatment for which the true value was included in the 95% confidence interval. The initial increase in CD4 during the first year was not correctly reproduced by Model 5. Models 6 and 7 which both incorporate the dynamics of viral load gave a better estimate of the increase in CD4 count (see details about estimated values of biological parameters in the Model 7 in Web-Supplementary Material A3). All models found a significant effect of treatment on CD4 counts in the first year. Sandwich estimators were used for calculating the standard errors for GEE methods and Fisher information matrix was used for methods based on maximum likelihood. Altogether, the dynamic models (Models 4–7) have greater statistical evidence, with higher Z-statistics rejecting the no-effect hypotheses, than the GEE-based models (Models 1–3). Finally, while fitting Models 1–6 took less than a minute on a typical laptop, fitting Model 7 took about 10 hours of parallel computing with 100 cores. All results and conclusions are similar in small and large samples.

4. Real Data

We used two large cohorts: the ANRS CO3 Aquitaine cohort (Thiébaut et al., 2000) and the Swiss HIV Cohort Study (SHCS) (Sterne et al., 2005; Gran et al., 2016). Like Cole et al. (2005), we took a sub-sample of patients who were alive, HIV positive, yet untreated and under follow-up in April 1996 when HAART became available. All patients taking only one or two antiretroviral drugs (rather than HAART) were excluded. Once a patient was on HAART, we assumed he or she remained on it. For each patient, follow-up began with the first visit after April 1996 and ended with 1) the last visit at which he or she was seen alive, 2) the last visit before patient discontinued the study, or 3) April 2003, whichever comes first. Data were assumed missing completely at random (MCAR); thus we deleted observations where either the viral load or CD4 count was missing. Patients with only one observation were excluded. After exclusions, there were 1591 patients from the Aquitaine cohort and 1726 patients from the SHCS (see Web-Supplementary Material B1 for a description of patient selection). Table 3 gives descriptive statistics. For most patients, follow up ended with administrative censoring and therefore we assumed censoring was not informative.

Table 4 displays the results we obtained for the effect of treatment on CD4 counts. The naive Model 1, not corrected for treatment assignment, indicated a small and non-significant increase in CD4 for SHCS cohort, and a significant negative effect for the Aquitaine Cohort; thus illustrating the need for modeling treatment assignment. This is corrected by the use of a treatment model in MSM Models 2. Models 3 and 4 led to similar results in the SHCS and different results in the Aquitaine cohort. For Model 4, this may be because covariates other than CD4 count are also confounders (such as viral load). For Model 3, some covariates driving the choice of treatment initiation may have been omitted in the treatment model. Estimates from both Models 3

Table 4

Estimated treatment effect on CD4 counts from real data of the Aquitaine cohort and SHCS: Model 1: naive regression; Model 2: MSM on Y_1^i; Model 3: MSM on Z_1^i; Model 4: simple LIM; Model 5: autoregressive LIM; Model 6: LIM system; Model 7: mechanistic model.

		Real data set observational studies					
		SHCS			Aquitaine cohort		
Model	β treatment	Effect	Sd.	Z-stat[a]	Effect	Sd.	Z-stat[a]
Model 1	<1 yr	6	16	0.34	−94	12	−7.55
	>1 yr	30	6	5.42	30	3	9.75
	∞	$+\infty$	–	–	$+\infty$	–	–
Model 2	<1 yr	208	18	11.31	36	19	1.87
	>1 yr	50	9	5.79	53	5	9.62
	∞	$+\infty$	–	–	$+\infty$	–	–
Model 3	<1 yr	174	10	17.29	27	14	1.99
	>1 yr	61	5	12.37	34	5	6.3
	∞	$+\infty$	–	–	$+\infty$	–	–
Model 4	<1 yr	189	11	17.33	109	9	12.03
	>1 yr	73	8	9.07	55	6	8.99
	∞	$+\infty$	–	–	$+\infty$	–	–
Model 5	<1 yr	26	*	*	45	*	*
	>1 yr	14	*	*	19	*	*
	∞	55	*	*	79	*	*
	β_1	60	4	16.04	14	3	4.12
Model 6	<1 yr	73	*	*	92	*	*
	>1 yr	26	*	*	16	*	*
	∞ CD4	104	*	*	111	*	*
	∞ VL	−2.0	*	*	−2.3	*	*
	β_1	80	16	5	28	18	1.53
	α_1	−3.29	0.09	−38.4	−3.19	0.1	−30.55
Model 7	<1 yr	104	*	*	71	*	*
	>1 yr	18	*	*	9	*	*
	∞ CD4	127	*	*	86	*	*
	∞ VL	−4.09	*	*	−3.14	*	*
	β	−1.73	0.05	−34.79	−0.89	0.01	−85.77

[a]Estimates for treatment effect (β) are significant at level 10% if the Z-stat is greater than 1.64 and significant at level 5% if the Z-stat is greater than 1.96.

*Simulated delta-method can lead to estimation of these values, but is not implemented here. Indeed, for Models 5, 6, and 7 significance of the treatment effect has to be evaluated through the mechanistic parameters β_1, α_1 and β.

and 4 suggest that treatment causes a significant increase in CD4 within the first year and in subsequent years. The 1-year increase, however, was much smaller for the Aquitaine cohort than for the SCHS when using MSM-based models. See Web-Supplementary Material B2 for a discussion of this result in relation with a possible practical violations of the experimental treatment assumption Cole and Hernán (2008). The results of the dynamical models, especially Models 6 and 7, were more consistent between cohorts.

Model 6 is interesting because it dissociates the effect of the treatment on CD4 count from its effects on viral load. In this model, the estimated treatment effect on CD4 count was small in both cohorts and was non-significant for the Aquitaine cohort. In contrast, the effect on viral load was highly significant in both cohorts. This is consistent with the mode of action of antiretroviral treatment: the increase in CD4 count is essentially mediated by the decrease in viral load; the latter is the direct effect of treatment. Such biological knowledge is incorporated in Model 7, where the treatment acts on the infectivity parameter. In view of the Z-statistics obtained by a Wald test of the hypothesis $\beta = 0$ in equation (9), the statistical evidence obtained in Model 7 is very high (this is confirmed by a likelihood ratio test). Moreover, Model 7 gives an insight into the value of the biological birth and death rates of cells during the infection (see Web-Supplementary Material B3). Regarding these biological parameters, the estimates from the two data sets are rather consistent in the sense that they have the same order of magnitude, although a formal comparison would show that several parameters are different, potentially due to different characteristics of the patients in the two cohorts. Finally, a simple way to look at these results and to compare them, is to consider the mean evolution in CD4 over time. Figure 2 represents the predicted CD4 counts with Models 1–7 for treated patients starting at baseline with CD4 count of 365 and a viral load of 4.4 (which are approximately the mean values at treatment initiation in these cohorts). For Models 1–3, these curves are deterministic, which is not the case for Models 4–7 because these models

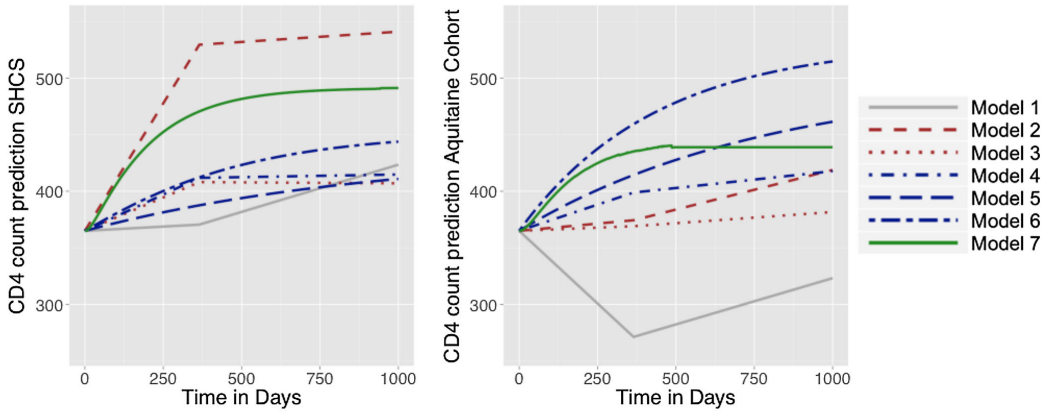


Figure 2. Mean evolution in CD4 predicted by Model 1 (plain line, simple regression), Model 2 (dashed line, MSM on Y_t^i), Model 3 (dotted line, MSM on Z_t^i), Model 4 (dotted line, simple LIM), Model 5 (dotted line, autoregressive LIM), Model 6 (dashed-dotted line, LIM system), and Model 7 (long-dashed line, mechanistic model) for treated patients starting with 365 CD4 cells/mL and a viral load of 4.4 log10 copies/mm^3: (left) estimates from the SHCS data (right) estimates from the Aquitaine cohort. This figure appears in color in the electronic version of this article.

have random effects. For these latter models, we computed the mean predicted curves depending on the value of the random effect, which have to be set to values compatible with the baseline values of the biomarkers. In order to set them, in both cases, we computed the equilibrium point of the system without treatment and solved the system of equations. Figure 2 shows that the naive Model 1 is not in agreement with medical knowledge. Models 1–3 are unstable, whereas Models 4–7 are more consistent between the two studies. Models 3, 4, and 5 lead to similar trajectories in the SHCS data set but very different trajectories in the Aquitaine Cohort study. In the latter, because trajectories for Model 3 look unrealistic and too pessimistic, we believe LIM Models 4 or 5 is more realistic. This illustrates the instability of weighted approaches. Finally, Model 5, 6, and 7 have an equilibrium point. Model 7 reaches a steady state in about 1 year which is in line with observed data and is similar in both cohorts.

5. Conclusion

In this article, we estimated the effect of HAART on CD4 count using four dynamic models and compared estimates with those from a naive regression model, a variant of a previously proposed MSM and a MSM based on linear increments. This is an empirical comparison rather than a theoretical comparison. We discussed assumptions and validity of the various approaches together with possible bias in estimates, statistical evidence, and practicability of use. The naive regression model (Model 1) strongly underestimated the effect of the treatment. The MSMs (Models 2 and 3) corrected this misleading result but sometimes failed to reach significance or were unstable across cohorts. The discrete-time dynamic models (Models 4, 5, and 6) based on LIM gave rather good estimates and show higher statistical evidence, although they may sometimes be too rigid. All the discrete-time models are easily fit without specialist software. The continuous-time dynamic model based on ODE-NLME (Model 7) gave good results. Models 6 and 7, which jointly model CD4 and viral load gave the most consistent results, with a richer interpretation since they explicitly model the effect of HAART on CD4 via a direct effect on viral load.

We have used a linear MSM with two slopes similar to that proposed by Cole et al. (2005). This model adequately represents the short-term (few years) effect of treatment but not the long-term effect because the effect in subsequent years implies a biologically implausible indefinite increase over time. As pointed out by the Associate Editor, because MSM can specify any reasonable outcome regression model, ranging from a very simple model which posits that only the most recent exposures affect the outcome to something very complex—e.g., a spline-weighted sequence of exposures (Xiao et al., 2014), it would be possible to define an MSM in which the effect would be bounded. However, this would be at the cost of additional non-linear parametrization. Also, it would be possible to use more recent methods such as the history adjusted MSM (Petersen et al., 2007); these are most suited for dynamic treatment regimes, whereas we assessed the effect of a static treatment regimes in this work. In contrast, most dynamical models (although not Model 4) have an equilibrium point. We showed that a MSM could be fitted for the increment Z_t^i rather than for Y_t^i. Thus, the MSM approach could complement the dynamic approach in the sense that less stringent assumptions would be needed for causal inference. However, the dynamic models already do a good job and the need to correct them (at the price of more complex procedure and loss of statistical evidence) is not obvious. In this article, we assumed MCAR observations for GEE, which is appropriate because most patients were administratively censored. MAR observations can be treated by using inverse probability of censorship weights (see Web-Supplementary Material Section B4). One advantage of likelihood-based dynamic models

is that they are still valid given MAR observations.

The mechanistic Model 7 directly incorporates biological knowledge. This leads to a more significant test for the parameter of interest. The simulated data in the article are obtained with a dynamic model, so one might think that this favors dynamic models; however, the true generation process comes itself from a complex dynamic system. Thus, the simulated model is much more complex than Models 4–7 which are highly misspecified. Of course, misspecification is not only a binary feature but can be quantified, in principle, for instance by the Kullback–Leibler risk between the generating distribution and the best distribution in the model (Commenges, 2015). Both simulated and real data are more complex than the models used to analyze that data: in both cases, we know that the model is misspecified, but we hope that the structure of the model captures essential features of the dynamics of the system. The comparison between dynamic and MSM approach remains empirical: we compare estimates from plausible MSM and dynamic models. Finally, mechanistic models, once estimated, open the possibility of designing optimal control of the therapy, as has been proposed on simulations by Adams et al. (2004), Ernst et al. (2006), and also Prague et al. (2012). The issue of "optimal treatment regime" has also been tackled outside of the context of mechanistic models (Petersen et al., 2007; Orellana et al., 2010; Saarela et al., 2015). The drawback of the continuous-time approach is that it is numerically challenging and requires special software running on cluster computers.

6. Web-Supplementary Materials

Web-Supplementary Material referenced in Sections 3, 4, and 5, the simulated data analyzed in Section 3 and a R program implementing Models 1–6 are available with this article at the *Biometrics* website on Wiley Online Library. Programs to estimate the parameters with Model 7 are available on a dedicated website: http://www.isped.u-bordeaux.fr/NIMROD.

Acknowledgements

The authors thank the investigators of the Aquitaine Cohort and the Swiss HIV Cohort Study. Parts of this work was founded by NIH grants R37 AI 51164. Parallel computing was used thanks to the MCIA (Mésocentre de Calcul Intensif Aquitain) of the Université de Bordeaux and of the Université de Pau et des Pays de l'Adour.

References

Aalen, O., Røysland, K., Gran, J., and Ledergerber, B. (2012). Causality, mediation and time: A dynamic viewpoint. *Journal of the Royal Statistical Society A* **175**, 831–861.

Adams, B., Banks, H., Davidian, M., Kwon, H., Tran, H., Wynne, S., et al. (2005). HIV dynamics: Modeling, data analysis, and optimal treatment protocols. *Journal of Computational and Applied Mathematics* **184**, 10–49.

Adams, B., Banks, H., Kwon, H.-D., and Tran, H. T. (2004). Dynamic multidrug therapies for HIV: Optimal and STI control approaches. *Mathematical Biosciences and Engineering* **1**, 223–241.

Arjas, E. and Parner, J. (2004). Causal reasoning from longitudinal data. *Scandinavian Journal of Statistics* **31**, 171–187.

Cole, S. and Hernán, M. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology* **168**, 656–664.

Cole, S., Hernán, M., Anastos, K., Jamieson, B., and Robins, J. (2007). Determining the effect of highly active antiretroviral therapy on changes in human immunodeficiency virus type-1 RNA viral load using a marginal structural left-censored mean model. *American Journal of Epidemiology* **166**, 219–227.

Cole, S., Hernán, M., Margolick, J., Cohen, M., and Robins, J. (2005). Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. *American Journal of Epidemiology* **162**, 471–478.

Commenges, D. (2015). Information theory and statistics: An overview. *arXiv:1511.00860*.

Commenges, D. and Gégout-Petit, A. (2009). A general dynamical statistical model with causal interpretation. *Journal of the Royal Statistical Society B* **71**, 719–736.

Commenges, D. and Gégout-Petit, A. (2015). The stochastic system approach for estimating dynamic treatments effect. *Lifetime Data Analalysis* **21**, 561–578.

Commenges, D. and Jacqmin-Gadda, H. (2015).*Dynamical Biostatistical Models*. New York, NY: CRC Press.

Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association* **95**, 407–424.

Didelez, V. (2008). Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society B* **70**, 245–264.

Diggle, P., Farewell, D., and Henderson, R. (2007). Analysis of longitudinal data with dropout: Objectives, assumptions and a proposal. *Journal of the Royal Statistical Society C* **56**, 499–550.

Eichler, M. and Didelez, V. (2010). On granger causality and the effect of interventions in time series. *Lifetime Data Analysis* **16**, 3–32

Ernst, D., Stan, G.-B., Goncalves, J., and Wehenkel, L. (2006). Clinical data based optimal STI strategies for HIV. In *Decision and Control, 2006 45th*, pages 667–672. San Diego: IEEE.

Gégout-Petit, A. and Commenges, D. (2010). A general definition of influence between stochastic processes. *Lifetime Data Analysis* **16**, 33–44.

Gran, J. M., Hoff, R., Roysland, K., Ledergerber, B., Young, J., and Aalen, O. (2016). Estimating the treatment effect of the treated under time-dependent confounding-applied to simulated data and to data from the swiss HIV cohort study. arXiv preprint arXiv:1604.01597.

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438.

Guedj, J., Thiébaut, R., and Commenges, D. (2007). Maximum likelihood estimation in dynamical models of HIV. *Biometrics* **63**, 1198–1206.

Hernán, M. A., Brumback, B., and Robins, J. (2002). Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Statistics in Medicine* **21**, 1689–709.

Hoff, R., Gran, J., and Farewell, D. (2014). Farewell's linear increments model for missing data: The FLIM package. *The R Journal* **6**, 137–150.

Ko, H., Hogan, J. W., and Mayer, K. H. (2003). Estimating causal treatment effects from longitudinal HIV natural history studies using marginal structural models. *Biometrics* **59**, 152–162.

Lavielle, M., Samson, A., Karina Fermin, A., and Mentré, F. (2011). Maximum likelihood estimation of long-term HIV dynamic models and antiviral response. *Biometrics* **67**, 250–259.

Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Orellana, L., Rotnitzky, A., and Robins, J. M. (2010). Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: Main content. *The International Journal of Biostatistics* **6**, 1557–4679.

Pepe, M. and Anderson, G. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics-Simulation* **23**, 939–951.

Perelson, A. (2002). Modelling viral and immune system dynamics. *Nature Reviews Immunology* **2**, 28–36.

Petersen, M. L., Deeks, S. G., Martin, J. N., and van der Laan, M. J. (2007). History-adjusted marginal structural models for estimating time-varying effect modification. *American Journal of Epidemiology* **166**, 985–993.

Petersen, M. L., Deeks, S. G., and van der Laan, M. J. (2007). Individualized treatment rules: Generating candidate clinical trials. *Statistics in Medicine* **26**, 4578–4601.

Petersen, M. L., Wang, Y., van der Laan, M. J., and Bangsberg, D. R. (2006). Assessing the effectiveness of antiretroviral adherence interventions: using marginal structural models to replicate the findings of randomized controlled trials. *Journal of Acquired Immune Deficiency Syndromes* **43**, S96–S103.

Prague, M., Commenges, D., Drylewicz, J., and Thiébaut, R. (2012). Treatment monitoring of HIV-infected patients based on mechanistic models. *Biometrics* **68**, 902–911.

Prague, M., Commenges, D., Guedj, J., Drylewicz, J., and Thiébaut, R. (2013). NIMROD: A program for inference via a normal approximation of the posterior in models with random effects based on ordinary differential equations. *Computer Methods and Programs in Biomedicine* **111**, 447–458.

Prague, M., Commenges, D., and Thiébaut, R. (2013). Dynamical models of biomarkers and clinical progression for personalized medicine: The HIV context. *Advanced Drug Delivery Reviews* **65**, 954–965.

Robins, J., Hernán, M., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560.

Saarela, O., Stephens, D. A., Moodie, E. E., and Klein, M. B. (2015). On bayesian estimation of marginal structural models. *Biometrics* **71**, 279–288.

Sterne, J., Hernán, M., Ledergerber, B., Tilling, K., Weber, R., Sendi, P., et al. (2005). Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: A prospective cohort study. *The Lancet* **366**, 378–384.

Thiébaut, R., Morlat, P., Jacqmin-Gadda, H., Neau, D., Mercié, P., Dabis, F., et al. (2000). Clinical progression of HIV-1 infection according to the viral response during the first year of antiretroviral treatment. *Journal of Acquired Immune Deficiency Syndromes* **14**, 971–978.

Walker, A. (1996). Confounding by indication. *Epidemiology* **7**, 335–336.

Wu, H. (2005). Statistical methods for HIV dynamic studies in AIDS clinical trials. *Statistical Methods in Medical Research* **14**, 171–192.

Xiao, Y., Abrahamowicz, M., Moodie, E. E., Weber, R., and Young, J. (2014). Flexible marginal structural models for estimating the cumulative effect of a time-dependent treatment on the hazard: reassessing the cardiovascular risks of didanosine treatment in the swiss HIV cohort study. *Journal of the American Statistical Association* **109**, 455–464.

Received July 2015. *Revised May* 2016. *Accepted June* 2016.

APPENDIX: Correspondence between parameters of Models 2, 3, and 4.

The question of how parameters from a MSM and a dynamical model relate is difficult for two reasons: the models are constructed differently; the philosophical approach to causality is different. We will make this exercise for comparing the MSM Model 2 and the LIM Model 4. One can reconcile the two philosophical approaches by saying that the "causal" interpretation (in a interventional point of view) is that for a given treatment trajectory $\overline{\boldsymbol{a}_t^i}$, we expect under the MSM: $\mathrm{E}(Y_t^i(\overline{\boldsymbol{a}_t^i})|Y_0^i) = \beta_0 + \beta_1 \mathrm{cum}(\overline{\boldsymbol{a}_{t-1}^i}) + \beta_2 \mathrm{cumlag}(\overline{\boldsymbol{a}_{t-1}^i}) + \beta_3 t + \beta_4 Y_0^i$. Model 4 is formulated in terms of the increments: $Z_t^i = \alpha_0 + \alpha_1 A_{t-1}^i + \alpha_2 A_{t-2}^i + b_i + \varepsilon_t^i$, which by summation gives $Y_t^i = Y_0^i + \alpha_1 \mathrm{cum}(\overline{\boldsymbol{A}_{t-1}^i}) + \alpha_2 \mathrm{cumlag}(\overline{\boldsymbol{A}_{t-1}^i}) + \alpha_3 t + M_t$, where $M_t^i = b_i + \sum_{k=1}^t \varepsilon_t^i$ is a martingale. This is the Doob decomposition of the process $\boldsymbol{Y}$. With the assumption of a "perfect" system or a NUC system (see Commenges and Gégout-Petit (2015) and Chapter 9 of Commenges and Jacqmin-Gadda (2015)), if we apply treatment trajectory $\overline{\boldsymbol{a}_t^i}$, this define a new probability P^a under which the Doob decomposition, which is: $Y_t^i = Y_0^i + \alpha_1 \mathrm{cum}(\overline{\boldsymbol{a}_{t-1}^i}) + \alpha_2 \mathrm{cumlag}(\overline{\boldsymbol{a}_{t-1}^i}) + \alpha_3 t + M_t^i$, from which we deduce: $\mathrm{E}(Y_t^i|Y_0^i) = Y_0^i + \alpha_1 \mathrm{cum}(\overline{\boldsymbol{a}_{t-1}^i}) + \alpha_2 \mathrm{cumlag}(\overline{\boldsymbol{a}_{t-1}^i}) + \alpha_3 t$. Thus, Model 4 yields the same expectation under an intervention imposing treatment trajectory $\overline{\boldsymbol{a}_{t-1}^i}$ as Models 2 if $\beta_4 = 1$, and the parameters giving the effects of $\mathrm{cum}(\overline{\boldsymbol{a}_{t-1}^i})$ and $\mathrm{cumlag}(\overline{\boldsymbol{a}_{t-1}^i})$ correspond. The same correspondence holds for Model 3, which is equivalent to Model 2 if the observations are equally spaced. Whereas Models 2 and 3 make the assumption that all confounders between $\boldsymbol{A}$ and $\boldsymbol{Z}$ are included in the computation of the inverse probability of treatment, Model 4 assumes that there is no confounder between $\boldsymbol{A}$ and $\boldsymbol{Z}$. For instance the viral load V_{t-1} might be a confounder which is taken into account in inverse probability of treatment in Models 2 and 3; in the dynamic approach one must use more complex models describing the dynamics of the viral load (such as Models 5 and 6). For Models 5, 6, and 7, marginal effects can still be computed (analytically or by simulation), but this may lead to complex forms, while generally MSM assume simple mathematical structures.

6.3 Prague et al. 2022 (CPT PsP) Model Building Strategy in Mechanistic Models

SAMBA: A novel method for fast automatic model building in nonlinear mixed-effects models. Prague M. and Lavielle M. *CPT: Pharmacometrics & Systems Pharmacology.* 11(2) - 161-172 - Feb 2022.

This article integrates in the first axis of my research "Statistical Methodology for Estimations and Building of Mechanistic Models". It has been written during my chaire at Ecole Polytechnique with Marc Lavielle, a collaborator from Inria team xpop. In 2023 and 2024, it received a "Top-Download Award," recognizing it as one of the top 10% most downloaded articles in its publication year.

I have selected this paper because it showcases a method that address a real problem in using mechanistic models for real-life modeling, i.e. building strategies. Indeed, because of computational complexity, it is required to deploy strategies that do not test all possible models but instead to select the one with the best information criteria. The main challenge was entering and modifying the code of the R package *Rsmlx* as well as finding good real-life examples to compare with existing methods.

Received: 5 July 2021 | Revised: 15 October 2021 | Accepted: 28 October 2021

DOI: 10.1002/psp4.12742

ARTICLE

SAMBA: A novel method for fast automatic model building in nonlinear mixed-effects models

Mélanie Prague[1,2] | **Marc Lavielle**[3]

[1]Inria Bordeaux Sud-Ouest, Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR 1219, University of Bordeaux, Bordeaux, France

[2]Vaccine Research Institute, Créteil, France

[3]Inria & CMAP, Ecole Polytechnique, CNRS, Institut Polytechnique de Paris, Paris, France

Correspondence
Mélanie Prague, ISPED - Université de Bordeaux - Bureau 23, 146 Rue Léo Saignat, 33070 Bordeaux Cedex, France.
Email: Melanie.Prague@inria.fr

Funding information
This study has received funding from the Nipah virus project financed by the French Ministry of Higher Education, Research, and Innovation

Abstract

The success of correctly identifying all the components of a nonlinear mixed-effects model is far from straightforward: it is a question of finding the best structural model, determining the type of relationship between covariates and individual parameters, detecting possible correlations between random effects, or also modeling residual errors. We present the Stochastic Approximation for Model Building Algorithm (SAMBA) procedure and show how this algorithm can be used to speed up this process of model building by identifying at each step how best to improve some of the model components. The principle of this algorithm basically consists in "learning something" about the "best model," even when a "poor model" is used to fit the data. A comparison study of the SAMBA procedure with Stepwise Covariate Modeling (SCM) and COnditional Sampling use for Stepwise Approach (COSSAC) show similar performances on several real data examples but with a much reduced computing time. This algorithm is now implemented in Monolix and in the R package *Rsmlx*.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
Existing model-building methods for nonlinear mixed-effects models have high computational time, especially for selecting the covariate model.

WHAT QUESTION DID THIS STUDY ADDRESS?
The study describes the principle of the Stochastic Approximation for Model Building Algorithm (SAMBA) procedure, which allows to build a covariate, a correlation, and an error model automatically and compares it with Stepwise Covariate Modeling (SCM) and COnditional Sampling use for Stepwise Approach (COSSAC) procedures.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
SAMBA allows to select the best covariate model without having to fit the complete nonlinear mixed-effects model to the data for each possible covariate model. This study confirms that it is possible to obtain relevant information on the model we are looking for, even when another model is fitted to the data. This allows to drastically reduce the computation time with respect to other existing procedures

while keeping the same performances. We also show that it is possible to perform correlation and error model selection in nonlinear mixed-effects models.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

This method will allow the practitioner to very quickly find a set of very good models in terms of data fitting and parsimony, even when the number of parameters or the number of covariates available is large.

INTRODUCTION

Construction of a complex (nonlinear) mixed-effects model[1] is a challenging process which requires confirmed expertise, advanced statistical methods, and the use of sophisticated software tools, but, above all, time and patience. Indeed, the success of correctly identifying all the components of the model is far from straightforward: it is a question of finding the best structural model, determining the type of relationship between covariates and individual parameters, detecting possible correlations between random effects, or also modeling residual errors. Our goal is to accelerate and optimize this process of model building by identifying at each step how best to improve some of the model components.

The procedure for constructing a model is usually iterative: one adjusts a first model to the data, and diagnosis plots and statistical tests allow to detect possible misspecifications in the proposed model. A new model must then be proposed to correct these defects and improve the predictive abilities of the model. Most of the common approaches consist in stepwise procedures consisting in testing the addition of variable forward and their elimination backward alternatively and progressing through the choice of models using a criterion derived from the log-likelihood. A widely used approach is Stepwise Covariate Modeling (SCM),[2] which consists in an exhaustive search in the covariates space. Each covariate addition or deletion is tested in turn selecting models at each step leading to the best adjustment according to the objective criterion. Approaches such as Wald Approximation Method (WAM)[3] and COnditional Sampling use for Stepwise Approach based on Correlation tests (COSSAC)[4] are less computationally intensive as they use, respectively, a likelihood ratio test and a correlation test to move in the covariates space, which allows the testing of less models. All these methods are nevertheless computationally intensive as they require to re-estimate the model parameters and the likelihood many times. In particular, these methods are very sensitive to "the curse of dimensionality" when the number of covariates to test on parameters is large.

The Generalized Additive Model (GAM) method[5,6] is computationally appealing as it does not require as many models fitting. Indeed, it is based on a regression on the empirical Bayes estimates (EBEs). The EBEs are the modes of the conditional distributions of the individual parameters. In other words, they are the most likely value of the individual parameters, given the estimated population parameters and the data. These estimates are known to be misleading and prone to shrinkage when data are sparse.[7] An efficient method which can correct the bias caused by the shrinkage of the EBEs have been recently proposed for covariate analysis.[8,9] In this paper, we propose to develop similar method which relies on the use of random samples from the conditional distribution of each individual parameters instead of EBEs. Indeed, the random sample of the posterior distribution has been shown to correctly control the type I error when performing tests to detect misspecifications in the model.[10]

As for most of the model-building procedures, the objective of Stochastic Approximation for Model Building Algorithm (SAMBA) is to find a model that minimizes some information criterion, such as Akaike information criterion (AIC), Bayesian Information Criteria (BIC), or corrected BIC (BICc).[11] The main principle of SAMBA is to use the results obtained with a wrong model to learn the right model. Then, SAMBA is an iterative procedure where a new model is used at each iteration of the algorithm. The values of the population parameters of the model are found by maximum likelihood estimation, and, then, the individual parameters are sampled from the conditional distribution defined under this estimated model. These simulated individual parameters combined with the observed data can now be used to select a new statistical model. It is important to underline that, as most of the iterative procedures for non-convex optimization, SAMBA does not pretend to be capable of always finding the global minimum of the used criterion, but it always allows to quickly find a very good solution.

Two contributions mainly constitute the content of this paper. First, we describe the novel algorithm called SAMBA for fast automatic model building in nonlinear mixed-effects models (section 1). Second, we benchmark its performances compared with reference methods SCM and COSSAC in real-world examples (section 2).

METHODS

Model description

Let $y_i = \left(y_{ij}, 1 \le j \le n_i\right)$ be the vector of observations for subject i, where $1 \le i \le N$. The model that describes the observations y_i is assumed to be a parametric probabilistic model that depends on a vector of L (individual) parameters $\psi_i = \left(\psi_{i1}, \ldots, \psi_{Li}\right)$. In a population framework, the vector of parameters ψ_i is assumed to be drawn from a population distribution $p(\psi_i)$. Then, defining a model $\mathscr{M}$ consists in defining a joint probability distribution for the observations $y = (y_1, \cdots, y_N)$ and for the individual parameters $\psi = (\psi_1, \cdots, \psi_N)$. For the sake of notation simplicity, we focus on models for continuous longitudinal data. However, extension to models for discrete data and time to event data is straightforward.

Let y_{ij}, the observation obtained from subject i at time t_{ij} be described as:

$$u\left(y_{ij}\right) = u\left(f\left(t_{ij}, \psi_i\right)\right) + g\left(t_{ij}, \psi_i, \xi\right)\epsilon_{ij}, 1 \le i \le N, 1 \le j \le n_i. \quad (1)$$

The structural model f is a fundamental component of the model because it defines the individual predictions of the observed kinetics for a given set of parameters. The residual errors (ϵ_{ij}) are assumed to be standardized Gaussian random variables (mean zero and variance 1). The residual error model is represented by function g in model (1) and may depends on some additional parameter ξ. Finally, one can use the function u to transform the observations, assuming for instance that they are log-normally distributed. In the following, we will assume u to be the identity.

We assume a linear model for the individual parameters (up to some transformation h):

$$h(\psi_i) = h(\psi_{\text{pop}}) + \beta c_i + \eta_i, \; 1 \le i \le N, \quad (2)$$

where $\eta_i \sim \mathcal{N}(0, \Omega)$ is a vector of random effects and where c_i is a vector of individual covariates used to explain part of the variability of the ψ_i's. The ψ_{pop} and β are fixed effects. The joint model of y and ψ then depends on a set of parameters $\theta = (\psi_{\text{pop}}, \beta, \Omega, \xi)$.

Selecting a model described by Equations 1 and 2 consists for the modeler in selecting: (i) the structural model f, (ii) the transformation of the individual parameters h, (iii) the residual error model g, (iv) the list of covariates that have an impact on individual parameters, and (v) the structure of the variance-covariance matrix of the random effects in the linear model Ω. The selection of the two first items is problem-specific, and their selection is out of the scope of this paper. We will therefore assume, in this paper, that f and h are given. The SAMBA procedure proposes solutions to address the selection of the three other components of the model.

The SAMBA procedure

Automatic model building is a difficult task because it is generally not possible to fit and compare all possible models. Moreover, it is necessary to define what is the "best model" among all the possible models. A classical approach consists in searching for the model $\mathscr{M}_*$, that minimizes a criterion, such as the penalized likelihood[12,13]:

$$\mathscr{M}_* = \arg\min_{\mathscr{M}} \left\{\min_\theta(-2\log(\mathscr{L}_{\mathscr{M}}(\theta; y))) + \text{pen}(\mathscr{M})\right\}. \quad (3)$$

The objective of this approach is to find a model that best fits the data (by minimizing $-2LL$) while being as simple as possible (it is the role of $\text{pen}(\mathscr{M})$ to favor models with few parameters). When the space of possible models is large, an exhaustive search is clearly impossible, and an efficient minimization strategy must be implemented. It is precisely for this purpose that SAMBA was developed: to obtain very quickly the "best" model $\mathscr{M}_*$, or a model with an objective criterion value very close to that of $\mathscr{M}_*$.

SAMBA is an iterative procedure alternating three steps. Assume that model $\mathscr{M}_k$ was obtained at iteration k of the algorithm. We first compute $\theta^{(k)}$, the maximum likelihood estimate of θ for model $\mathscr{M}_k$. We then generate a set of individual parameters $\psi^{(k)}$ from the conditional distribution of individual parameters $p_{\mathscr{M}_k}(\psi | y; \theta^{(k)})$. The selection step finally consists in building a new model $\mathscr{M}_{k+1}$ using the *complete data* $(y; \psi^{(k)})$ and minimizing the complete penalized criterion:

$$\mathscr{M}_{k+1} = \arg\min_{\mathscr{M}} \left\{\min_\theta\left(-2\log\left(\mathscr{L}_{\mathscr{M}}\left(\theta; y, \psi^{(k)}\right)\right)\right) + \text{pen}(\mathscr{M})\right\}. \quad (4)$$

As already mentioned, the statistical model to be built consists of a covariate model, a correlation model, and a residual error model. Then, the selection of model $\mathscr{M}_{k+1}$ is composed of three model selection procedures: the selection of the covariate model $\mathscr{M}_{k+1}^{\text{COV}}$, the selection of the correlation model $\mathscr{M}_{k+1}^{\text{CORR}}$, and the selection of the error model $\mathscr{M}_{k+1}^{\text{ERR}}$. Note that not all these components are necessarily selected: some may have been set arbitrarily because of existing knowledge. By noticing that $\mathscr{L}_{\mathscr{M}}\left(\theta; y, \psi^{(k)}\right) = \mathscr{L}_{\mathscr{M}}(\theta | y, \psi^{(k)})\mathscr{L}_{\mathscr{M}}\left(y, \psi^{(k)}\right)$, it appears that the problem of selecting the error model is independent from the problem of selecting the covariate and correlation models. Figure 1 provides a flowchart of the complete procedure. Let us now take a closer look at what each step of the model selection process consists of.

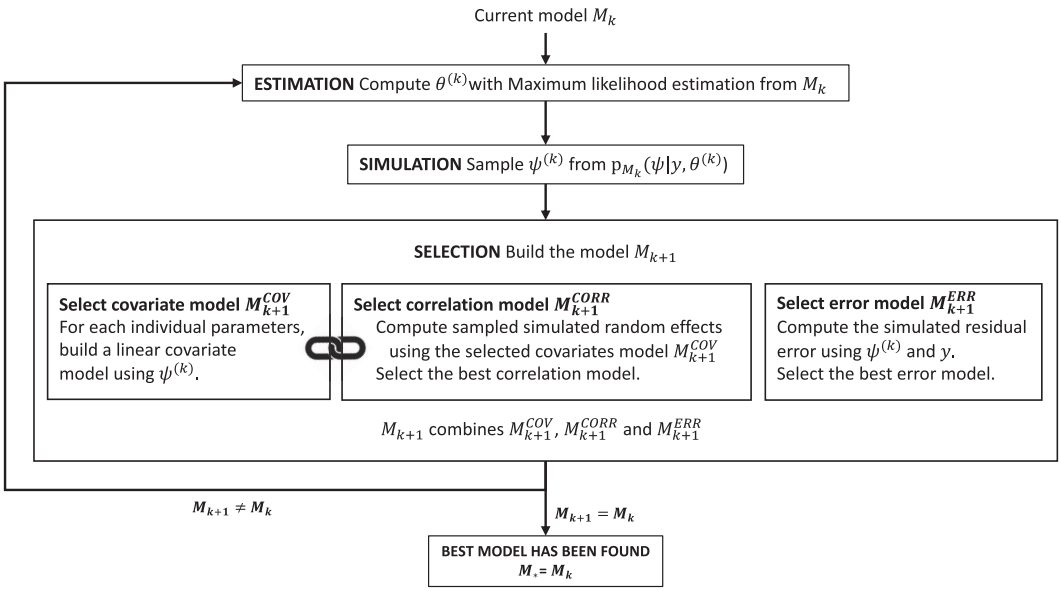


FIGURE 1 Scheme of the Stochastic Approximation for Model Building Algorithm (SAMBA)

The covariate model selection $\mathscr{M}_{\mathbf{k+1}}^{\mathbf{COV}}$

The sample $\psi^{(k)}$ has been generated conditionally to the data y and the model $\mathscr{M}_k$. For the ℓ-th parameter, we build a linear model between $\psi_\ell^{(k)}$ and covariates c, such as in Equation 2:

$$h_\ell\left(\psi_{i\ell}^{(k)}\right) = h_\ell\left(\psi_{\text{pop},\ell}\right) + \beta_\ell c_i + \eta_{i\ell}^{(k)},\ 1 \le i \le N, 1 \le \ell \le L, \tag{5}$$

with h_ℓ the transformation associated to the ℓ-th parameter and where $\eta_{i\ell}^{(k)}$ is supposed normally distributed with mean zero and variance ω_ℓ^2. We define $\theta_\ell = \left(\psi_{\text{pop},\ell}, \beta_\ell, \omega_\ell^2\right)$. Best covariate model for parameter ℓ, denoted $\mathscr{M}_{k+1}^{\text{COV}_\ell}$, is selected as being the one minimizing a penalized criterion:

$$\mathscr{M}_{k+1}^{COV_\ell} = \arg\min_{\mathcal{M}}\left\{\min_{\theta_\ell}\left(-2\log\left(\mathscr{L}_{\mathscr{M}}\left(\theta_\ell;\psi_\ell^{(k)}\right)\right)\right) + \text{pen}^{\text{COV}}(\mathscr{M})\right\}.$$

We denote n_β the number of non-null elements in β_ℓ for model $\mathscr{M}$. The penalization depends on the criterion selected for optimization: if AIC then $\text{pen}^{\text{COV}}(\mathscr{M}) = 2n_\beta$, if BIC or BICc then $\text{pen}^{\text{COV}}(\mathscr{M}) = \log(N)\, n_\beta$. Equation 5 tells us that the covariate selection problem has become here a classical problem of variable selection in a linear model.[14] This problem is much more easily tractable than the original one. The overall best covariate model combines the best model for each parameter such that $\mathscr{M}_{k+1}^{\text{COV}} = \left\{\mathscr{M}_{k+1}^{\text{COV}_1}, \ldots \mathscr{M}_{k+1}^{\text{COV}_L}\right\}$.

In the implemented version of package *Rsmlx* (R speaks Monolix), two different strategies are implemented depending on the dimension of the selection problem. If the number d of available covariates is less than 11, an exhaustive search is performed over all the 2^d possible covariate models for each parameter. Otherwise, the stepwise variable selection procedure implemented in the function *stepAIC* from package *MASS* is used. It consists of iteratively adding and removing covariates in stepwise manner to lower the objective criterion.

The correlation model selection $\mathscr{M}_{\mathbf{k+1}}^{\mathbf{CORR}}$

Using the selected covariate model $\mathscr{M}_{k+1}^{\text{COV}}$ and the sample of individual parameters $\psi_i^{(k)}$, it is possible to extract the vector of individual random effects $\eta_i^{(k)} = \left(\eta_{i\ell}^{(k)}, \ell = 1, \ldots, L\right)$ from Equation 5. Assuming that $\eta_i^{(k)} \sim \mathcal{N}(0, \Omega)$ where Ω is a block diagonal matrix, the problem of correlation model selection consists in selecting the block structure of Ω. We then select the correlation model denoted $\mathscr{M}_{k+1}^{\text{CORR}}$ by minimizing a penalized criterion:

$$\mathscr{M}_{k+1}^{\text{CORR}} = \arg\min_{\mathscr{M}}\left\{\min_{\Omega}\left(-2\log\left(\mathscr{L}_{\mathscr{M}}\left(\Omega;\eta_i^{(k)}\right)\right)\right) + \text{pen}^{\text{CORR}}(\mathscr{M})\right\}.$$

We denote n_Ω the number of non-zero elements in the upper triangular part of the matrix Ω. The penalization

depends on the criterion selected for global optimization: if AIC then $\text{pen}^{\text{CORR}}(\mathscr{M}) = 2n_\Omega$, if BIC or BICc then $\text{pen}^{\text{CORR}}(\mathscr{M}) = \log(N)\, n_\Omega$.

In the implemented version of package *Rsmlx*, we limit the size of the block-structure that can be considered at each iteration. For $\mathscr{M}_1$, no correlation can be added and a diagonal matrix is used for Ω; for $\mathscr{M}_2$ only blocks of size two are considered. At iteration k for selection of model $\mathscr{M}_{k+1}^{\text{CORR}}$, block size cannot be larger than $k+1$, leading to no more than $(k-1)k/2$ non-zero covariance terms in Ω.

The error model selection $\mathscr{M}_{\mathbf{k+1}}^{\mathbf{ERR}}$

For a given set of simulated individual parameters $(\psi_i^{(k)}, 1 \le i \le N)$, the residual errors can easily be computed:

$$e_{ij}^{(k)} = y_{ij} - f\left(t_{ij}, \psi_i^{(k)}\right), 1 \le i \le N,\ 1 \le j \le n_i.$$

We then fit several error models with standard deviation of the form $g\left(t_{ij}, \psi_i^{(k)}, \xi\right)$ for $e_{ij}^{(k)}$ and select the one minimizing a penalized criterion:

$$\mathscr{M}_{k+1}^{ERR} = \text{argmin}_{\mathscr{M}}\left\{\min_\xi\left(-2\log\left(\mathscr{L}_{\mathscr{M}}\left(\xi; e_{ij}^{(k)}\right)\right)\right) + \text{pen}^{\text{ERR}}(\mathscr{M})\right\}.$$

We denote n_ξ the length of ξ (i.e., the number of parameters in model $\mathscr{M}$). The penalization depends on the criterion selected for global optimization: if AIC then $\text{pen}^{\text{ERR}}(\mathscr{M}) = 2n_\xi$, if BIC then $\text{pen}^{\text{ERR}}(\mathscr{M}) = \log(N)\, n_\xi$, and if BICc then $\text{pen}(\mathscr{M}) = \log\left(n_{\text{tot}}\right) n_\xi$ where n_{tot} is the total number of observations, including below the limit of quantification data.

In the implemented version of package *Rsmlx*, five error models (provided by function gin Equation 1) are tested by default: constant $\left(g_x\left(t_{ij}, \psi_i^{(k)}, \xi\right) = \xi\right)$, proportional $\left(g_x\left(t_{ij}, \psi_i^{(k)}, \xi\right) = \xi f\left(t_{ij}, \psi_i\right)\right)$, combined_1 $\left(g_x\left(t_{ij}, \psi_i^{(k)}, \xi\right) = \xi_1 + \xi_2 f\left(t_{ij}, \psi_i\right)\right)$, combined_2 $\left(g_x\left(t_{ij}, \psi_i^{(k)}, \xi\right) = \sqrt{\xi_1^2 + \xi_2^2 f\left(t_{ij}, \psi_i\right)}\right)$, or exponential in which a constant error model is fitted to the $\log(y)$ using the transformation $u = \log$ in Equation 1. Note that it is currently not possible to perform the selection on a restricted number of error models, but such a feature could be easily implemented.

Stopping rule procedure

At each iteration k of the algorithm, we combine $\mathscr{M}_{k+1}^{\text{COV}}$, $\mathscr{M}_{k+1}^{\text{CORR}}$, and $\mathscr{M}_{k+1}^{\text{ERR}}$ to get the new selected model $\mathscr{M}_{k+1}$, which is passed forward on to the next estimation-simulation run. It is important to select the covariate model before the correlation model. On the other hand, the error model can be updated before or after the other two components of the model. The algorithm stops when $\mathscr{M}_k$ is strictly identical to $\mathscr{M}_{k+1}$ for all components and the last model is the selected one.

Remark

In the above, $\psi_i^{(k)}$ represents a single realization of the conditional distribution $p_{\mathscr{M}_k}(\psi_i|y, \theta^{(k)})$ for each $i = 1, \ldots N$. Instead of considering only one realization of this distribution, we could use a sample of size $R\left(\psi_{i\ell,r}^{(k)}, 1 \le r \le R\right)$. If so, the linear covariate model described in Equation 5 rewrites:

$$\overline{h_\ell\left(\psi_{\ell,i}^{(k)}\right)} = h_\ell\left(\psi_{\ell,\text{pop}}\right) + \beta_\ell c_i + \overline{\eta_{\ell,i}^{(k)}},\ 1 \le i \le N,\ 1 \le \ell \le L,$$

where:

$$\overline{h_\ell\left(\psi_{\ell,i}^{(k)}\right)} = \frac{1}{R}\sum_{r=1}^{R} h_\ell\left(\psi_{i\ell,r}^{(k)}\right)$$

Procedures for covariate model selection and correlation model selection remains the same, but using now $\left(\overline{\psi_{i\ell}^{(k)}}\right)$ and $\left(\overline{\eta_{i\ell}^{(k)}}\right)$ at iteration k. On the other hand, the R series of residual errors $\left(e_{ij,r}^{(k)}\right)$ are used for selecting the residual error model.

RESULTS

Step-by-step example of the SAMBA procedure

To illustrate how SAMBA works in practice, we will describe step-by-step the complete procedure on the example of remifentanil.[15] We use here the SAMBA implementation in function *buildmlx* of the R package *Rsmlx*, using the default settings.

The remifentanil data

The dataset is composed of 65 healthy adults who have received remifentanil i.v. infusion at a constant infusion rate between 1 and 8 $\mu g^{-1}\, kg^{-1}\, min^{-1}$ for 4 to 20 minutes. Time and rate of infusion are known for each individual. The pharmacokinetic (PK) data consists in the plasma concentration of remifentanil, which is measured during and after infusion for a total of 19 to 53 observations by patients, totaling 2057 observations. A total of six

covariates are available: one qualitative covariate, the sex (SEX) and five continuous covariates: the age (AGE), the height (HT), the weight (WT), the lean body mass (LBM), and the body surface area (BSA). All the latter are normalized and log-transformed for the analysis. In the following, we adopt the notation $\text{logAGE} = \log\left(\text{AGE}/\text{AGE}_{\text{pop}}\right)$, where AGE_{pop} is a typical value to normalize on (e.g., the mean value of age in the population).

The model

The PK model for i.v. infusion has a central compartment (volume V1), two peripheral compartments (volumes V2 and V3, and intercompartmental clearances Q2 and Q3), and a linear elimination (Cl). Log-normal distributions are used for the six individual parameters. The $2^6 = 64$ possible covariate models will be considered for each of the six individual parameters. Note that if we had to test all possible models, we would have had to test 64^6 combinations, which would have made the problem intractable.

SAMBA iterations

We start the SAMBA procedure with a model $\mathscr{M}_0$ without any covariate on all parameters, with no correlation between random effects and the so-called combined_1 error model. Figure 2 illustrates the selection steps on this specific example. One can notice that the BICc, which has been chosen as target criterion, decreases from 7186 for $\mathscr{M}_0$ to 6985 for $\mathscr{M}_1$, 6957 for $\mathscr{M}_2$, and 6903 for $\mathscr{M}_3$, which is finally selected as the best model for this example.

- **Run 0 (BICc = 7185.8) + Iteration 1:** Model $\mathscr{M}_0$ is fitted to data and individual parameters are sampled conditionally on the data and this model. Each of the 64 possible linear covariate models is fitted to each individual parameters and the one with lowest BICc is selected. Let us take the example of *Cl*: the three best models include (1) an effect of logAGE and logWT (BICc = −55.0), (2) an effect of logAGE and logLBM (BICc = −56.1), and (3) an effect of logAGE and logBSA (BICc = −57.5). The latter is chosen as the best model for parameter *Cl* as it provides the lowest BICc ($\mathscr{M}_1^{\text{COV},Cl}$). Altogether, for all parameters, the best covariate model ($\mathscr{M}_1^{\text{COV}}$) includes logAGE on all parameters, logBSA on *Cl*, and logLBM on *V*1 and *V*2. No correlation is added to the model because no correlation is allowed at first iteration. Then, $\mathscr{M}_1^{\text{CORR}}$ is a diagonal variance-covariance matrix for the random effects. Among the tested error models, the three best ones are proportional (BICc = 5815.2), combined_1 (BICc = 5811.2), and combined_2 (BICc = 5807.0), which is selected for $\mathscr{M}_1^{\text{ERR}}$. These covariate, correlation, and error models are then passed on to run 1: $\mathscr{M}_1 = \{\{\mathscr{M}_1^{\text{COV},Cl}, \mathscr{M}_1^{\text{COV},Q2}, \mathscr{M}_1^{\text{COV},Q3}, \mathscr{M}_1^{\text{COV},V1}, \mathscr{M}_1^{\text{COV},V2}, \mathscr{M}_1^{\text{COV},V3}\}, \mathscr{M}_1^{\text{CORR}}, \mathscr{M}_1^{\text{ERR}}\}\}$.
- **Run 1 (BICc = 6984.9) + Iteration 2:** Model $\mathscr{M}_1$ is fitted to the data and individual parameters are sampled. Again, the three best model for each covariate are provided. The best covariate model includes logAGE on all parameters except *V*1, logBSA on *Cl*, logLBM on *V*1, and SEX on *V*2 ($\mathscr{M}_2^{\text{COV}}$). Block-structured correlation with blocks up to size 2 are compared (i.e., up to one correlation term). The best three models are with a correlation between parameters *Cl* and *V*2 (BICc = 1082.9), between parameters *Cl* and *Q*2 (BICc = 1093.8), and between parameters *V*2 and *Q*2 (BICc = 1072.0). The latter correlation model is selected for $\mathscr{M}_2^{\text{CORR}}$. Residual error model combined_2 remains the best one ($\mathscr{M}_2^{\text{ERR}}$). These covariate, correlation, and error models are then passed on to run 2.
- **Run 2 (BICc = 6956.9) + Iteration 3:** Model $\mathscr{M}_2$ is fitted to data and individual parameters are sampled. The best covariate model includes logAGE on all parameters except *V*1, logBSA on *Cl*, and logLBM on *V*1 and *V*2 ($\mathscr{M}_3^{\text{COV}}$). Block-structured correlation with blocks up to size 3 are compared (i.e., up to three correlation terms), a correlation block is selected between *Cl*, *Q*2, and *V*2 ($\mathscr{M}_3^{\text{CORR}}$). Residual error model combined_2 remains the best one ($\mathscr{M}_3^{\text{ERR}}$). These covariate, correlation, and error models are then passed on to run 3.
- **Run 3 (BICc = 6903.4) + Iteration 4:** Model $\mathscr{M}_3$ is fitted to data and individual parameters are sampled. Of note, regarding the correlation model selection, block-structured correlation with blocks up to size 4 are compared (i.e., up to six correlation terms). During this iteration, the same model as the one in the previous iteration is selected ($\mathscr{M}_4 = \mathscr{M}_3$) resulting in the stopping of the procedure. Model $\mathscr{M}_3$ is therefore the final model selected with the SAMBA procedure.

Converging toward a global optimal model

Even if the selected criterion decreases at each iteration, there is no guarantee that SAMBA converges toward a global minimum of this criterion. The quality and the robustness of the convergence of SAMBA can then be assessed by running SAMBA several times from different starting models. In particular, a good practice is to: (1) launch SAMBA from several initial models, (2) compare the best models found (if there is not only one) in terms of objective criterion (e.g., BICc), and (3) make

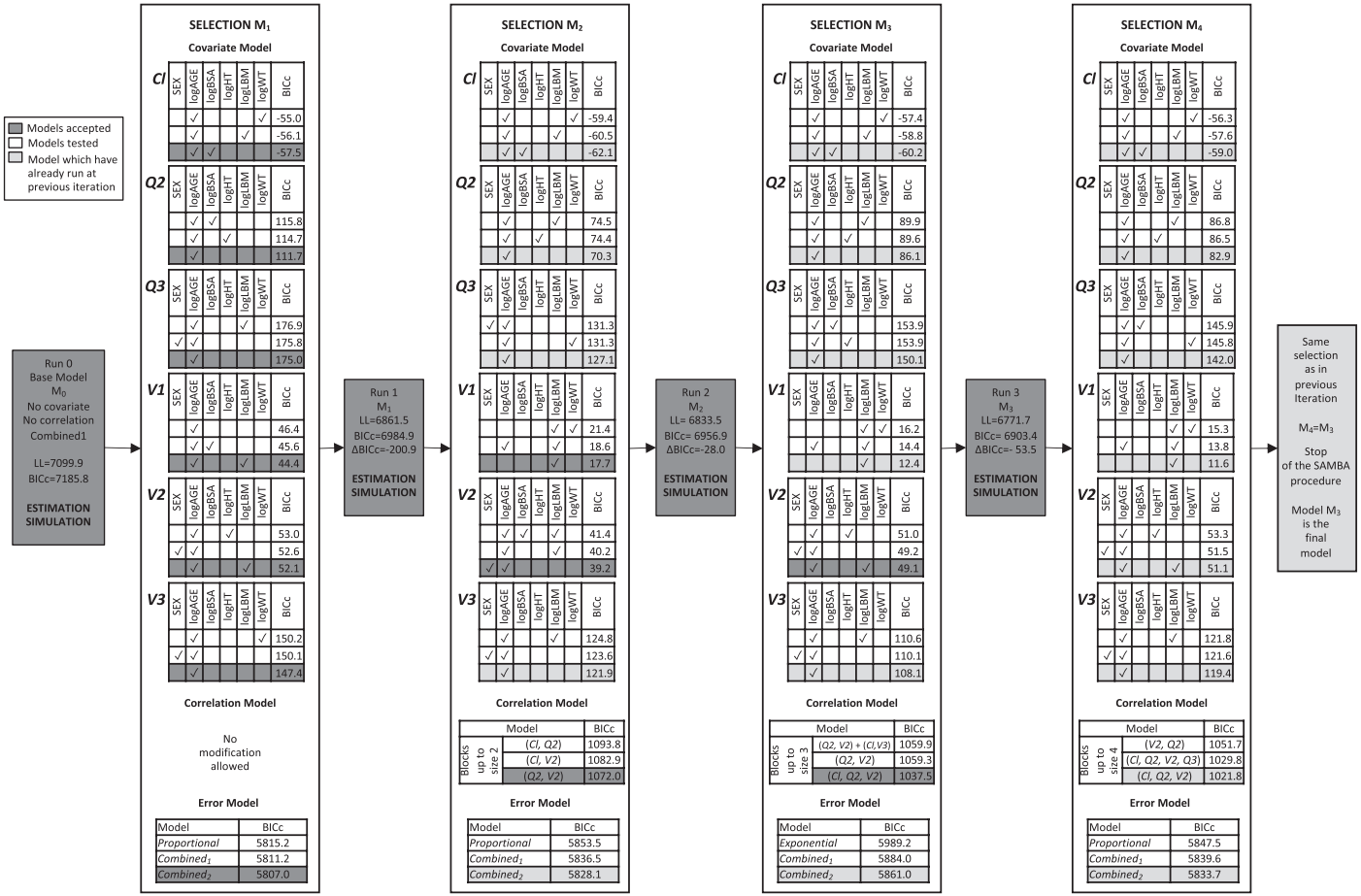


FIGURE 2 2 Step-by-step Stochastic Approximation for Model Building Algorithm (SAMBA) procedure on the remifentanil example with six covariates (SEX, logAGE, logBSA, logHT, logLBM, and logWT) and six model parameters (*Cl*, *Q2*, *Q3*, *V1*, *V2* and *V3*). For each selection (covariate, correlation, and error model), the three best models in term of corrected Bayesian Information Criteria (BICc) are displayed. Non selected models are in white, newly accepted models are in darker grey, and models which have been already accepted at previous run are in lighter grey

a thorough analysis and interpretation of the nearby models in order to choose the most relevant one for a given application. Regarding the choice of the starting model, similarly to the Expectation Maximization and Stochastic Approximation Expectation Maximization algorithms, there is no optimal choice.[16,17] We recommend to test in priority the following three starting models: (1) an empty model, (2) (when possible) a complete model, and (3) a model (or models) that make sense for the biological application. Note that this robustness assessment is standard for all non-convex optimization algorithms and should also be performed for SCM and COSSAC in routine.

Performances on real examples, and comparison with the SCM and COSSAC procedures

To assess the performances of the SAMBA procedure compared to SCM and COSSAC procedures, we replicate the illustration provided in ref. 4. We applied the three routines to a collection of 10 representative datasets, including PKs, pharmacodynamics, and disease models. Of note, the SCM method for variable selection used here is exactly the same as the one implemented in PsN (Pearl Speaks NONMEM), differences lie in the algorithms used to estimate the parameters of a model and to calculate the likelihood. We restricted the SAMBA procedure to the covariate model selection as correlation and error model selection are not implemented in COSSAC and SCM. The results can be found in Table 1.

Because the datasets are real data illustrations, there is no "true" model. It is only possible to compare them in terms of BIC. Of 10 examples, the same best model was proposed by the three procedures in four examples. In two examples, the best model selected by SAMBA was better in terms of BICc than with SCM and COSSAC (Theophylline Ext. Rel. and Warfarin PK/PD). In three other examples, the model with the lowest BICc was not selected by SAMBA. However, the difference in BICc was, respectively, smaller than six in comparison with the SCM procedure and 4.2 in comparison with the COSSAC procedure. We insist on the fact that a difference in BICc does not necessarily have any biological meaning. This is an arbitrary criterion that allows to quantify the goodness of fit with respect to the sparsity of the model chosen. We thus argue that the three procedures lead to rather similar models, which all constitute very good starting points for the modeler to build a model based on biological hypothesis. Finally, in only one example discussed below, the difference in BICc was larger than 10 points of BICc both compared with the SCM and COSSAC procedures.

Regarding the cholesterol dataset, we again ran the SAMBA procedure starting from a full model in which all covariates are supposed to have an effect on all parameters. The new model selected by SAMBA is the full model with an effect of logAGE on (*Chol*0, *slope*) and SEX on (*Chol*0, *slope*) is much closer in term of BICc than the one selected starting from an empty model (Δ BICc $= -2$). We can finally notice with this example that it is sometimes possible to improve the convergence of SAMBA by improving the convergence of SAEM. Indeed, using 10 Markov chains instead of only one, SAMBA also finds the model selected by SCM and COSSAC. Finding the optimal settings that minimizes computation time while maximizing the probability of finding the best model is an extremely difficult problem that remains open. We can claim that the default settings used in Rsmlx and Monolix give very good results in most cases, but not in all cases with absolute certainty.

In terms of computational effort, it is important to note that the SAMBA procedure completes the model-building process in much less runs, hence much less CPU time than SCM and COSSAC. In the considered problems, the number of runs and the CPU computation time are equivalent because the other computation times are negligible in the order of a few seconds. Actually, the computation times are six to 149 smaller than for SCM and two to 11 times smaller than for COSSAC. Note that the number of evaluations required by SAMBA is always lower or equal to the number of evaluations performed by COSSAC and SCM.

Simulation study

Data generation and analysis

We simulated data from a one-compartment PK model. The model has three population parameters $\mathrm{ka}_{\mathrm{pop}} = 1$, $V_{\mathrm{pop}} = 10$ and $Cl_{\mathrm{pop}} = 2$. All individual parameters are log-normally distributed around the population parameters ($\omega_{ka} = 0.2$, $\omega_V = 0.3$ and $\omega_{Cl} = 0.3$). We simulated five individual covariates $(C_1, C_2, C_3, C_4, C_5)$ from standard normal distributions. The covariate model is such that there only exists linear relationships between $\log(V)$ and C_1 ($\beta_{V,1} = 0.2$), $\log(Cl)$ and C_1 ($\beta_{Cl,1} = -0.2$), and $\log(Cl)$ and C_2 ($\beta_{Cl,2} = 0.3$). The correlation model is such that there exists a linear correlation between η_V and η_{Cl} ($\rho_{V,Cl} = 0.6$). Finally, the error model is a combined$_2$ model with $a = 2$ and $b = 0.1$. A clinical trial could then be simulated by generating PK data from this model for 100 individuals and 11 timepoints (0.25, 0.5, 1, 2, 5, 8, 12, 16, 20, 24, and 30). In order to evaluate the properties of SAMBA by Monte-Carlo, we simulated 100 replicates of the same trial and built the model for each replicate using SAMBA as implemented in *Rsmlx* and Monolix for minimizing BICc.

TABLE 1 Comparison of the SAMBA procedure with the SCM and COSSAC procedure on 10 representative datasets

		SCM		COSSAC		SAMBA		ΔBICc	
Dataset	**Characteristics**	**#Runs[b]**	**Final Model[a]**	**#Runs[b]**	**Final Model 1**	**#Runs[b]**	**Final Model[a]**	**SAMBA-SCM**	**SAMBA-COSSAC**
Warfarin	32 ind. - 247 obs.	44	logtWt - *V*, *Cl*	4	Identical	2	Identical	0	0
Linear PK	4 param. - 3 cov.		logtAge - *C*						
	4 re - 1 outcome								
Remifentanil	65 ind. - 1992 obs.	295	logLBM – *V*1	13	logLBM - *V*1, ***V***2	4	logLBM - *V*1	0.8	0.5
Linear PK	6 param. - 6 cov.		logAGE - *Cl*, Q2, **Q**3, *V* 2, *V* 3		logAGE - *Cl*, *Q*2, *V* 2, *V* 3		logAGE - *Cl*, *Q*2, **Q**3, *V* 2, *V* 3		
	4 re - 1 outcome		logBSA - *Cl*		logBSA - *Cl*		logBSA - *Cl*		
			logHT - **V**2						
					SEX - **V**3		SEX - **V**2		
Theophylline	12 ind. - 20 obs.	12	logtWEIGHT - **ka**	4	Identical	2	Identical	0	0
Linear PK	3 param. - 2 cov.								
	4 re - 1 outcome								
Quinidine	136 ind. - 361 obs.	22	none	11	Identical	1	Identical	0	0
Sparse PK	3 param. - 2 cov.								
	3 re - 1 outcome								
Tobramycin	97 ind. - 322 obs.	22	logCLCR - *Cl*	6	logCLCR - *Cl*	2	logCLCR - *Cl*	4.2	4.2
Sparse PK	3 param. - 2 cov.		logWT - **V**		logWT - **V**		logWT - **Cl**		
	2 re - 1 outcome								
Theophylline	18 ind. - 362 obs.	98	logWT - **Tlag**1, **V**	8	logWT - **Tlag**1	6	logWT - **F**, **V**	−11.7	−27
Ext. Rel.	7 param. - 3 cov.				logAGE -**ka**2		logAGE -**F**		
Linear PK	7 re - 1 outcome						logHT **ka**1, **ka**2, **Tlag**1, **diffTlag**2		
Warfarin	32 ind. - 247+232 obs.	92	logWT - *Cl*	10	logWT - *Cl*	2	logWT - *Cl*, **V**	−1.4	−1.4
PK/PD	8 param. - 3 cov.						logAGE -**Cl**, **R**0		
Joint	8 re - 2 outcomes								
Cholesterol	200 ind. - 1044 obs.	12	logAGE - *Chol*0, **slope**	5	logAGE - *Chol*0, **slope**	2	logAGE - *Chol*0	13.5	13.5
Disease	2 param. - 2 cov.		SEX - **slope**		SEX - **slope**				
Progression	2 re - 1 outcome								
Alzheimer	896 ind. - 3707 obs.	73	APOE - *alpha*, *p*0	8	APOE - *alpha*, *p*0	2	APOE - *alpha*, *p*0	6	1.5

(Continues)

TABLE 1 (Continued)

Dataset	Characteristics	SCM		COSSAC		SAMBA		ΔBICc	
		#Runs[b]	Final Model[a]	#Runs[b]	Final Model 1	#Runs[b]	Final Model[a]	SAMBA-SCM	SAMBA-COSSAC
Sparse PK	2 param. - 7 cov.		logAGE - *p0*, **alpha**		logAGE - *p0*, **alpha**		logAGE - *p0*		
	2 re - 1 outcome		logBMI - **alpha**		logBMI -**alpha**				
			logWT - *p0*		logWT - *p0*		logWT - *p0*		
Tranexamic	166 ind. - 817 obs.	298	GROUP - *Cl*, *V2*	12	Identical	2	Identical	0	0
PK	4 param. - 10 cov.		logBMI - Cl						
	4 re - 1 outcome		logCOCK - Cl						
			logLBW - Q						
			logWeight - V2						

[a]Differences of variable selection between different methods are highlighted in bold.
[b]The number of runs is defined as the number of time the estimation and the simulation steps are performed (which is the most time-consuming).

TABLE 2 Performance of the SAMBA algorithm for the selection of the covariate model in a simulation study using a one-compartment PK model

	Rsmlx			Monolix		
Covariates	***ka***	***V***	***Cl***	***ka***	***V***	***Cl***
C_1	2	100	100	2	100	100
C_2	0	1	100	0	1	100
C_3	1	2	1	2	2	1
C_4	0	3	4	0	3	4
C_5	0	1	1	1	2	1

One hundred datasets of 100 individuals with 11 observations each have been generated. True model $\mathscr{M}^*$ includes an effect of C_1 on V and Cl and an effect of C_2 on Cl. The percentages of times (over 100 replicates) each covariate-parameter relationship is selected in the final model are displayed. Implementation of SAMBA in *Rsmlx* and Monolix are compared.

Abbreviations: Cl, linear elimination; ka, absorption rate constant; PK, pharmacokinetic; SAMBA, Stochastic Approximation for Model Building Algorithm; V, volume.

The initial model did not include any covariate-parameter relationship and any correlation between random effect. The initial residual error model was a combined$_1$ model. The R code used for this Monte-Carlo study is available as Supplementary Material.

Performances

Table 2 summarizes the results obtained for the covariate model selection. On the one hand, we can see that, for this particular example, SAMBA finds the three existing covariate-parameter relationships in 100% of the cases. On the other hand, very few spurious relationships are detected (less than 2%). Importantly, in all cases for which the final covariate model included more covariates than the true model M^*, the BICc of the selected model was lower than that of M^* (the differences ranging from 3 to 14.7 with *Rsmlx* and from 2.4 to 14.6 for Monolix). In other words, SAMBA always finds a covariate model as good or better than M^* in terms of BICc. Regarding the selection of the correlation model, the correct model was selected for all the replicates. Finally, the correct error model was selected in 86% of the times with *Rsmlx* and 85% of the times with Monolix. Note that all the wrong selected error models were all combined$_1$ model (instead of combined$_2$) with a slightly larger BICc most of the time. Actually, these two models are quite similar and difficult to distinguish on the basis of a criterion like BICc. SAMBA then may get stuck in a local minimum in such a situation. Finally, and importantly, the final selected models obtained with *Rsmlx* and Monolix are different in only 6% of cases. These small differences are due to small

differences in the implementation of the algorithm (see the Discussion section for more details).

DISCUSSION

This paper presents a novel model-building procedure which offers covariate, correlation, and error model selection. It is fast as it requires only a limited number of runs of population parameter estimation and simulation compared to SCM and COSSAC. It allows to explore the space of models rapidly and provides to the modeler a very good model in term of the selection criterion. However, we insist on the fact that this procedure does not aim at replacing model-building based on biological knowledge, which is, in essence, the strength of mechanistic modeling. Thus, it should not be blindly used and the best—potentially few best—models should be interpreted and compared.

SAMBA is an efficient algorithm for minimizing an objective function. In this paper, we do not aim at evaluating the quality of the criterion used for model selection.[18] What is of interest here is the convergence of SAMBA. As it is also the case for SCM and COSSAC, SAMBA may not converge to the global minimum. This is particularly the case when the amount of data is too small compared to the complexity of the model to build. This phenomenon will be particularly critical when the number of covariates is high and/or when these are highly correlated. We then strongly encourage the user to build strategies to assess the robustness of the results. Extensions of the proposed algorithm are possible but are outside the scope of this paper and constitute a possible new research direction.

When there is a large number of available covariates, COSSAC and mainly SCM often fail in finding the best model in a reasonable time. In this case, SAMBA represents a particularly appealing approach because the covariate model selection is based on a stepwise variable selection procedure for linear models, which is known to handle high-dimension problems. Although stepwise AIC/BIC are designed to obtain a sparse estimator that works well on the training set, other methods, such as the lasso,[19] where the penalty is chosen with cross validation, is designed to obtain the sparse linear model that minimize the prediction error. A lasso type approach[20] can sometimes present better performances than an approach based on an information criterion, such as AIC or BIC, in particular when the number of covariates is very high. However, it should be noted that the choice of the penalty parameter by cross-validation can be complicated to implement and require a large number of runs. This type of method could be alternatively implemented in the covariate selection procedure and compared in further works. Note finally that it would be interesting to study the behavior of SAMBA using the EBEs (corrected as proposed in ref. 8,9), rather than the individual simulated parameters, to build the covariate model.

The SAMBA procedure is implemented the R Package *Rsmlx* in the function *buildmlx*.[21] Minimal required input is a Monolix project used as initial model. Additional arguments can be used to enable specific features (all not listed): select the components of the model to optimize among the covariate, correlation, and error model, restrict the number of parameters or covariates to use, select a specific objective criterion, etc. *Rsmlx* is on CRAN and the R code can be modified to investigate any of the alternative implementations mentioned above for a specific problem. Note that the execution of *Rsmlx* requires the Monolix software, because it is only an algorithm combining tasks implemented in Monolix. The R codes allowing to replicate the analyses of this paper are available in the Supplementary Material. All the illustration datasets can be downloaded from the Supporting Information Appendix S2 of ref. 4.

Finally, the SAMBA procedure is also implemented in the Monolix-GUI software starting from version 2019. Implementation is similar to the one in *Rsmlx* with two noteworthy differences. First, for the selection of covariates, a stepwise procedure is used even if the number of covariates d is small. Second, compiling differences exist between C++ and R. The full SAMBA procedure is available in the model-building perspective, under a task called automatic statistical model building method. A single iteration of the SAMBA procedure is also proposed in the section Proposal in the tab Results after running a single estimation and simulation step for a model in Monolix.[22]

CONFLICT OF INTEREST

Marc Lavielle is chief scientist of Lixoft, the company that develops and distributes the Monolix Suite. The other author declared no competing interests for this work.

AUTHOR CONTRIBUTION

M.P. and M.L. wrote the manuscript, designed the research, performed the research, and analyzed the data. M.L. contributed new reagents/analytical tools.

REFERENCES

1. Lavielle M. *Mixed-effects Models for the Population Approach: Models, Tasks, Methods and Tools*. CRC Press; 2014.
2. Jonsson EN, Karlsson MO. Automated covariate model building within NONMEM. *Pharm Res*. 1998;15(9):1463-1468.
3. Kowalski KG, Hutmacher MM. Efficient screening of covariates in population models using Wald's approximation to the likelihood ratio test. *J Pharmacokinet Pharmacodyn*. 2001;28(3):253-275.
4. Ayral G, Si Abdallah J-F, Magnard C, Chauvin J. A novel method based on unbiased correlations tests for covariate selection in

nonlinear mixed-effects models–the COSSAC approach. *CPT Pharmacomentrics Sys Pharmacol*. 2021;10(4):318-329.
5. Hastie T, Tibshirani R. Generalized additive models: some applications. *J Am Stat Assoc*. 1987;82(398):371-386.
6. Mandema JW, Verotta D, Sheiner LB. Building population pharmacokinetic/pharmacodynamic models. I. models for covariate effects. *J Pharmacokinet Biopharm*. 1992;20(5):511-528.
7. Nguyen T, Mouksassi M-S, Holford N, et al. Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT: Pharmacom Sys Pharmacol*. 2017;6(2):87-109.
8. Yuan M, Xu XS, Yang Y, et al. A quick and accurate method for the estimation of covariate effects based on empirical bayes estimates in mixed-effects modeling: correction of bias due to shrinkage. *Stat Methods Med Res*. 2019;28(12):3568-3578.
9. Yuan M, Zhu Z, Yang Y, et al. Efficient algorithms for covariate analysis with dynamic data using nonlinear mixed-effects model. *Stat Methods Med Res*. 2021;30(1):233-243.
10. Lavielle M, Ribba B. Enhanced method for diagnosing pharmacometric models: random sampling from conditional distributions. *Pharm Res*. 2016;33(12):2979-2988.
11. Delattre M, Lavielle M, Poursat M-A. A note on BIC in mixed-effects models. *Elect J Stat*. 2014;8(1):456-475.
12. Green PJ. Penalized likelihood. *Encyclopedia Stat Sci*. 1998;2:578-586.
13. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning, volume 112*. Springer; 2013.
14. George EI. The variable selection problem. *J Am Stat Assoc*. 2000;95(452):1304-1308.
15. Minto CF, Schnider TW, Egan TD, et al. Influence of age and gender on the pharmacokinetics and pharmacodynamics of remifentanil: I. model development. *J Am Soc Anesthesiol*. 1997;86(1):10-23.
16. Baudry J-P, Celeux G. EM for mixtures. *Stat Comp*. 2015;25(4):713-726.
17. Biernacki C, Celeux G, Govaert G. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate gaussian mixture models. *Comput Stat Data Anal*. 2003;41(3–4):561-575.
18. Buatois S, Ueckert S, Frey N, Retout S & Mentré F. Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models. *The AAPS J*. 2018;20(3):1-9.
19. Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc: Ser B (Methodol)*. 1996;58(1):267-288.
20. Hastie T, Tibshirani R, Tibshirani R. Best subset, forward stepwise or lasso? analysis and recommendations based on extensive comparisons. *Stat Sci*. 2020;35(4):579-592.
21. Lavielle M. Rsmlx: R Speaks 'Monolix'. 2021. http://rsmlx.webpopix.org. R package version 3.0.3.
22. Monolix Online Documentation - Proposal Tab Description. https://monolix.lixoft.com/tasks/proposal/. Accessed September 15, 2021.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Prague M, Lavielle M. SAMBA: A novel method for fast automatic model building in nonlinear mixed-effects models. *CPT Pharmacometrics Syst Pharmacol*. 2021;00:1-12. doi:10.1002/psp4.12742

6.4 Alexandre et al. 2021 (SMMR) Analysis of HIV Antiretroviral Interruption Trials

Between-group comparison of area under the curve in clinical trials with censored follow-up: Application to HIV therapeutic vaccines. Alexandre M., **Prague M.**, Thiébaut R. *Statistical Methods in Medical Research.* 30(9) - 2130-2147 - July 2021.

This article integrates in the second axis of my research "Within-host Modeling in Infectious Diseases". It has been written by a PhD student that I co-directed.

I chose this paper because it tackles a crucial challenge in HIV vaccine development: assessing treatment efficacy. While not primarily focused on mechanistic models (though extensions are currently being considered, including using model averaging), it employs spline models to manage informative loss of follow-up, aligning with a MAR mechanism. A significant challenge was deriving the analytical formulation for the mean and variance of the time-averaged area under the viral load curve.

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Between-group comparison of area under the curve in clinical trials with censored follow-up: Application to HIV therapeutic vaccines

Statistical Methods in Medical Research
2021, Vol. 30(9) 2130–2147

DOI: 10.1177/09622802211023963
journals.sagepub.com/home/smm
SAGE

Marie Alexandre[1,2], Mélanie Prague[1,2] and Rodolphe Thiébaut[1,2]

Abstract

In clinical trials, longitudinal data are commonly analyzed and compared between groups using a single summary statistic such as area under the outcome versus time curve (AUC). However, incomplete data, arising from censoring due to a limit of detection or missing data, can bias these analyses. In this article, we present a statistical test based on splines-based mixed-model accounting for both the censoring and missingness mechanisms in the AUC estimation. Inferential properties of the proposed method were evaluated and compared to ad hoc approaches and to a non-parametric method through a simulation study based on two-armed trial where trajectories and the proportion of missing data were varied. Simulation results highlight that our approach has significant advantages over the other methods. A real working example from two HIV therapeutic vaccine trials is presented to illustrate the applicability of our approach.

Keywords

Area under the curve, longitudinal data, statistical test, mixed-effects model, study drop out, left-censoring

1 Introduction

The area under the curve (AUC) is a summary measure commonly used in various applications when the outcome of interest is based on a quantitative variable such as a biomarker concentration. In pharmacokinetics, the AUC of the drug concentration versus time is typically analyzed to account for drug exposure and clearance from the body[1] or to evaluate the bioequivalence of vaccines,[2] or the quality of life by summarizing individual scores.[3–6] In preclinical cancer drug screening tumor xenograft experiments, the ratio or the difference of AUC can be used to replace the commonly used treatment-to-control ratio[7,8] or summarize symptoms[9] to evaluate therapy effectiveness. In infectious diseases, the AUC can summarize the exposure to the HIV virus[10] or influenza.[11,12] When AUC is an outcome to be compared between arms in a clinical trial, estimates can be biased because of incomplete data. Two frequent sources for the lack of completeness can arise: censoring due to a limit of detection (LOD) of assay and study drop out.

In this context, various methods for the calculation of AUC have been proposed. Allisson et al.[13] and Venter et al.[14] compared different approaches based on incremental AUC. Incremental AUC consists in computing the AUC only for observations that are above a threshold, which can be viewed as particularly compelling when there

[1]University of Bordeaux, Inria Bordeaux Sud-Ouest, Inserm, Bordeaux Population Health Research Center, SISTM Team, France
[2]Data Science Division, Vaccine Research Institute (VRI), Créteil, France

Corresponding author:
Mélanie Prague, University of Bordeaux, Inria Bordeaux Sud-Ouest, Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR 1219, F-33000 Bordeaux, France.
Email: melanie.prague@inria.fr

is left-censored observations. However, Potteiger et al.[15] pointed out the potential bias in resulting conclusions when using incremental AUC even in presence of complete data. Wilding et al.[16] have developed an approach to evaluate treatment effect by comparing longitudinal data from two groups of patients through AUC calculation when data are subject to missing completely at random (MCAR) missingness process. Bell et al.[17] extended this method to missing at random (MAR) data and incorporated the within-subject variability through random effects using linear mixed effects models (LMEMs). In both cases, the comparison of the mean AUC using maximum likelihood (ML) between groups was more robust than the comparison of the average individuals' AUC with standard two-sample *t*-tests. Furthermore, the estimation of the mean AUC using LMEM can be adapted to outcomes subject to left-censoring.[18]

In this paper, we propose a statistical parametric test for AUC based on splines-based MEMs which is extending the previously described approaches by adding flexibility in the modeling, accounting for left-censored data and dealing with MAR monotonic censored follow-up. Estimation of parameters in LMEMs model is possible using ML-based approach leading to robust inference in presence of right-censored[19] and left-censored outcome.[20,21] To do so, we use an expectation-maximization EM algorithm for computing the maximum likelihood in nonlinear mixed effects models with censored response as describe in Vaida et al.[22]

Multiple other non-parametric approaches have been developed to solve this type of problem. Schisterman and Rotnizky[23] developed a semi-parametric estimator of a K-sample U-statistic when data are missing at random combining information from both outcomes and auxiliary variables. Thereafter, Spritzler et al.[24] extended these results by proposing a valid semi-parametric two-sample test of equal AUC when observations are MAR monotonic and/or missing completely at random (MCAR). Both works are based on weighting approaches and thus require strong assumptions on the missing data process. Alternative non parametric tests have been developed by Vardi et al.[25] based on permutation tests. However, parametric approaches may help in the situation of incomplete data.

This work was motivated by the evaluation of HIV therapeutic vaccine in clinical trials where high rate of censoring can occur. The goal of the vaccines in HIV-infected patients is to boost the immune system to control the viral replication when antiretroviral treatments (ART) are interrupted. Hence, analytical treatment interruption (ATI) is the ultimate way to assess the ability of new vaccine strategies to control viral replication after ART discontinuation.[26] However, HIV-infected patients undergoing ATIs are subject to high risks of immune damage with expansion of the existing reservoir, clinical symptoms, resistance emergence, increased risk of HIV transmission as well as loss of therapeutic benefits from ART.[27,28] Therefore, ATI periods are short and patients are followed carefully. Specification of criteria determining ART resumption may vary from one study to another: development of Grade-3 adverse events or AIDS-related events, the CD4 cell count fell below 350 cells/mm^3, or a HIV RNA load exceeding a given virologic threshold.[29–34] Following these criteria, ART resumption may occur before the end of the planned ATI period leading to missing data comparable to study drop out. Also, HIV RNA viral load is subject to left censoring due to LOD usually around 50 copies/mL.[20] Therefore, the comparison of AUC in HIV therapeutic vaccine trials constitutes a particularly relevant context for the application of the method described in the paper.

The article is structured as follows. In section 2, we briefly describe two HIV therapeutic vaccine studies which motivated the development of our ML based-model proposed approach to estimate the difference of mean AUCs between two groups of patients when observations are left-censored and subject to follow-up censoring presented in section 3. In section 4, we investigate the inferential properties of this method and compare them with both traditional methods and a non-parametric test through simulation studies. To illustrate the applicability of the approach, we provide a real working example from the two motivating examples in section 5. To conclude, we summarize the paper and propose future research in section 6.

2 Motivating examples

In this paper, we focus on two HIV therapeutic vaccine trials testing the efficacy of vaccines through ART interruption in HIV-1-infected patients. The first one is the HIV therapeutic vaccine trial VRI02 ANRS 149 LIGHT.[35] This study is a randomized double-blind, two-arm placebo-controlled Phase-II trial. Its primary objective was to evaluate the virological efficacy after ART interruption of a therapeutic immunization compared to a placebo. The therapeutic immunization is based on a recombinant DNA vaccine (GTU-MultiHIV B) and a lipopeptide vaccine (LIPO-5). This study enrolled 105 patients (35 in the placebo control group vs. 70 in the

vaccinated group) whose 91 of them (32 placebo and 59 vaccinated) experienced ATI. HIV RNA load was repeatedly measured at times 0, 2, 4, 6, 8 and 12 weeks after ATI. The second study is the HIV therapeutic vaccine trial ANRS 093 Vac-IL2 (Vac-IL2).[36] This study is a randomized two-arm placebo-controlled Phase-II trial enrolling 71 patients (37 in the control group and 34 in the vaccinated group). Its primary objective was to evaluate the immunogenicity of a therapeutic immunization strategy combining two different vaccines, recombinant ALVAC-HIV (vCP1433) and Lipo-6T (HIV-1 lipopeptides), followed by the administration of subcutaneous interleukin-2 (IL-2). Therapeutic immunization was followed by 12 weeks of ATI with repeated measures of HIV RNA load at times 0, 1, 2, 3, 4, 6, 8, 10, 12 weeks after ATI.

3 Method

3.1 Definition of the AUC by interpolation method

We consider N subjects divided into G vaccine arms, with $N = \sum_{g=1}^{G} n_g$, with n_g being the number of patient in group g. Let $Y_{ij,g}$ be the response measured for the subject i belonging to group g at its jth time point, $t_{ij,g}$, with $i \in \{1, \cdots, N\}, j \in \{1, \cdots, m_i\}$ and $g \in \{1, \cdots, G\}$. Moreover, we define $\{t_{ij,g}\}$ as the set of time points at which data are observed for the patient i and $m_i = |\{t_{ij,g}\}|$ the cardinal of this set. At group level, we equivalently note $\{t_{j,g}\} = \cup_{i \in g}(\{t_{ij,g}\})$ the set of time points at which outcome of interest is measured for at least one patient in g, whose m_g is the cardinal. As defined, this framework allows the consideration of unbalanced group design and group-specific time points. The area under the response of interest curve can be calculated by the trapezoid interpolation method. The AUC summary measure for the ith subject belonging to the group g and summary statistics for the entire group g can then be approximated by the following equations. Without loss of generality, we define the lower limit of the integration interval as well as the first time point in each group as zero

$$\mathrm{AUC}_i = \int_0^{T_i} Y_{i,g}(t)\mathrm{d}t \simeq \sum_{j=2}^{m_i} \frac{(t_{ij,g} - t_{ij-1,g})}{2}(Y_{ij,g} + Y_{ij-1,g})$$

$$\mathrm{AUC}_g = \int_0^{T_g} \overline{Y}_g(t)\mathrm{d}t \simeq \sum_{j=2}^{m_g} \frac{(t_{j,g} - t_{j-1,g})}{2}(\overline{Y}_{j,g} + \overline{Y}_{j-1,g})$$

where $\overline{Y}_{j,g}$ is defined as the mean value of the outcome Y in the gth group at its jth time point, $\overline{Y}_{j,g} = \frac{1}{n_g}\sum_{i \in g} Y_{ij,g}$, $T_i = \max_j(\{t_{ij,g}\})$ and $T_g = \max_j(\{t_{j,g}\})$ the individual and group time of follow-up. Whereas the trapezoid method is known as the cumulative area over $m - 1$ time period in which the value of interest Y is approximated by a straight line between two adjacent points (t_{j-1}, y_{j-1}) and (t_j, y_j), two other interpolation methods have been studied in this work to approximate AUC using either global or piecewise cubic polynomials instead of linear function: (1) the Lagrange method and (2) the Spline method (see Online Appendices A and B for more details, respectively). These methods are not described in the main body of the article as they provide similar results to the described trapezoid interpolation method.

When calculating individual's AUC, it is usual to divide the AUC by the delay of follow-up to take into account the variability in follow-up due to early drop-out for example.[37–40] Although we propose in this article a method based on modeling that would allow to work directly on the raw AUC, we will use a normalized AUC (nAUC), that is the AUC divided by the number of days/weeks of follow-up, for the sake of comparison with individual level methods. The nAUC are given by equations (1) and (2)

$$n\mathrm{AUC}_i = \frac{1}{T_i}\int_0^{T_i} Y_{i,g}(t)\mathrm{d}t \simeq \frac{1}{T_i}\sum_{j=2}^{m_i} \frac{(t_{ij,g} - t_{ij-1,g})}{2}(Y_{ij,g} + Y_{ij-1,g}) \quad (1)$$

$$n\mathrm{AUC}_g = \frac{1}{T_g}\int_0^{T_g} \overline{Y}_g(t)\mathrm{d}t \simeq \frac{1}{T_g}\sum_{j=2}^{m_g} \frac{(t_{j,g} - t_{j-1,g})}{2}(\overline{Y}_{j,g} + \overline{Y}_{j-1,g}) \quad (2)$$

3.2 Estimation of nAUC by mixed effects model

We assume the MEM given by equation (3) to describe the outcome $Y_{ij,g}$ of the subject i in the group g_i at the jth time point

$$Y_{ij,g_i} = f_0(t_{ij,g_i}) + \sum_{g=1}^{G} \mathbb{1}_{[g_i=g]} \times F_g(t_{ij,g}) + h_i(t_{ij,g_i}) + \varepsilon_{ij} \tag{3}$$

where the function f_0 gathers all non-group-specific terms, e.g. an intercept, the functions F_g are non-linear smooth functions of time describing the fixed effect specific to each group and h_i are polynomial time-dependent random effects modeling the inter-individual variability. In the following, the functions F_g are set to linear combinations such as $F_g(t_{ij,g}) = \sum_{k=1}^{K_g} \beta_k^g f_g^k(t_{ij,g})$ where K_g is the number of time-dependent components describing the group-specific dynamics, e.g. spline basis, and β_k^g are the regression coefficients.

For generalization purpose, the LMEM given in equation (3) can be re-expressed with matrix formulation as follow

$$\boldsymbol{Y} = \boldsymbol{X_0}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{Y}$ is the vector of the outcome of interest, $\boldsymbol{X_0}$, $\boldsymbol{X}$, and $\boldsymbol{Z}$ are respectively the design matrices for the non-group- and group-specific fixed effects and random effects. Because vaccine or randomized controlled trials involve often adjustment of treatment effects on covariates, such as baseline covariates, the use of MEM allows it through the definition of the design matrices, whether at population, group or individual level. The vectors $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$ and $\boldsymbol{b}$ are the unknown non group- and group-specific fixed parameters and the random parameters respectively, while $\boldsymbol{\varepsilon}$ is the vector of error terms supposedly normally distributed such as $\mathbb{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{\varepsilon}) = \Theta$. Moreover, we assume that $\mathbb{E}(\boldsymbol{b}) = \boldsymbol{0}$ and $\mathrm{Var}(\boldsymbol{b}) = \Omega$, with $\boldsymbol{b} \perp\!\!\!\perp \boldsymbol{\varepsilon}$. By construction, the matrix $\boldsymbol{X}$ is defined as a diagonal block matrix such as $\boldsymbol{X} = \mathrm{diag}(\boldsymbol{X_1}, \cdots, \boldsymbol{X_G})$, where each sub-matrix $\boldsymbol{X_g}$ is group-specific. Similarly, the vector $\boldsymbol{\beta}$ can be written as $\boldsymbol{\beta^T} = \left(\boldsymbol{\beta^{1^T}}, \cdots, \boldsymbol{\beta^{G^T}}\right)$, each vector $\boldsymbol{\beta^g}$ being only specific to the group g. It can be demonstrated that the estimate of the nAUC in group g (2) can be re-expressed as a linear combination of the responses at each time, as

$$n\mathrm{AUC}_g = \frac{1}{T_g} \sum_{j=1}^{m_g} w_{j,g} \overline{Y}_{j,g} = \frac{1}{T_g} \boldsymbol{w_g}^T \overline{\boldsymbol{Y}}_{\boldsymbol{g}} \tag{4}$$

where $\boldsymbol{w_g} = (w_{1,g}, \cdots, w_{m_g,g})^T$, $\overline{\boldsymbol{Y}}_{\boldsymbol{g}} = (\overline{Y}_{1,g}, \cdots, \overline{Y}_{m_g,g})^T$, with

$$w_{j,g} = \begin{cases} \dfrac{t_{j+1,g} - t_{j,g}}{2}, & j = 1 \\ \dfrac{t_{j,g} - t_{j-1,g}}{2}, & j = m_g \\ \dfrac{t_{j+1,g} - t_{j-1,g}}{2}, & \text{otherwise} \end{cases} \tag{5}$$

In our method, the approximation of the summary statistics nAUC is obtained post-estimation of the MEM parameters. To this end, we denote $\widehat{\boldsymbol{\mu}}_{\boldsymbol{g}} = \mathbb{E}(\widehat{\boldsymbol{Y}}_{\boldsymbol{g}})$ being the expected value of the estimation of $\boldsymbol{Y}$ in the gth group, where $\widehat{\boldsymbol{\mu}}_{\boldsymbol{g}} = (\widehat{\mu}_{1,g}, \cdots, \widehat{\mu}_{m_g,g})^T$ with $\widehat{\mu}_{j,g} = \mathbb{E}(\widehat{Y}_{j,g})$ and $\widehat{\boldsymbol{Y}}_{\boldsymbol{g}} = (\widehat{Y}_{1,g}, \cdots, \widehat{Y}_{m_g,g})^T$. It follows that $\widehat{\mu}_{j,g}$ is expressed as a linear combination of the fixed parameter estimates denoted $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$ for the group- and non-group-specific. Indeed, by noting $X_0^{[g]}$ the sub-matrix of $\boldsymbol{X_0}$ corresponding to the group g, we obtain $\widehat{\boldsymbol{\mu}}_{\boldsymbol{g}} = \boldsymbol{X}_0^{[g]}\widehat{\boldsymbol{\gamma}} + \boldsymbol{X_g}\widehat{\boldsymbol{\beta}}^{\boldsymbol{g}}$ leading to

$$\widehat{\mu}_{j,}g = \sum_{v=1}^{\dim(\widehat{\gamma})} X_{0jv}^{[g]} \cdot \widehat{\gamma}_v + \sum_{v=1}^{K_g} X_{gjv} \cdot \widehat{\beta}_v^g$$

Replacing $\overline{\boldsymbol{Y}}_{\boldsymbol{g}}$ by $\widehat{\boldsymbol{\mu}}_{\boldsymbol{g}}$ in equation (4), the approximation of nAUC in the group g, $\widehat{n\mathrm{AUC}}_g$, can be written as

$$\widehat{n\mathrm{AUC}}_g = \frac{1}{T_g} \boldsymbol{w_g^T} \widehat{\boldsymbol{\mu}}_g \tag{6}$$

3.3 Statistical testing of difference between groups

We want to identify whether or not two groups of treatment can be differentiated by their mean value of the area under the response curve. Consequently, we defined the hypotheses of interest for the two compared groups g and $\tilde{g}$ as the equality and the difference of their nAUC for the null hypothesis, H_0 and the alternative one, H_1, respectively.

While the mechanism of follow-up censoring and the resulting missing data have no direct impact on the method of the MEM estimation, the statistical test must be written to take it into account. The presence of informative censoring impacting directly the time of follow-up and thus the time interval of AUC calculation for each group, $[0, T_g]$, the statistical test is build to compare the mean value of AUC on the same time interval. To do this, we define the upper integration limit for nAUC calculation as $T = \min(T_g, T_{\tilde{g}})$ given the time restricted nAUC for each group calculated as

$$n\widehat{\mathrm{AUC}}_g^{\mathrm{rest}} = \frac{1}{T}\int_0^T \widehat{\mu}_g(t)\mathrm{d}t \simeq \frac{1}{T}\mathring{\boldsymbol{\omega}}_g^T \widehat{\boldsymbol{\mu}}_g^{\mathrm{rest}} \tag{7}$$

where $\mathring{\boldsymbol{\omega}}_g = (\omega_{1,g}, \cdots, \omega_{\mathring{m}_g,g})^T$ and $\widehat{\boldsymbol{\mu}}_g^{\mathrm{rest}} = (\widehat{\mu}_{1,g}, \cdots, \widehat{\mu}_{\mathring{m}_g,g})^T$ with $\mathring{m}_g = |\{t_{j,g} | t_{j,g} \leq T\}|$.

Based on equation (7) of the approximation of nAUC in the group g, the test hypotheses may be re-expressed in terms of model fixed parameters such as

$$\mathrm{H}_0 : n\widehat{\mathrm{AUC}}_g^{\mathrm{rest}} = n\widehat{\mathrm{AUC}}_{\tilde{g}}^{\mathrm{rest}} \iff \frac{1}{T}\ \mathring{\boldsymbol{\omega}}_g^T\left(\mathring{\boldsymbol{X}}_0^{[g]}\widehat{\boldsymbol{\gamma}} + \ \mathring{\boldsymbol{X}}_g\widehat{\boldsymbol{\beta}}^{\boldsymbol{g}}\right) = \frac{1}{T}\ \mathring{\boldsymbol{\omega}}_{\tilde{g}}^T\left(\mathring{\boldsymbol{X}}_0^{[\tilde{g}]}\widehat{\boldsymbol{\gamma}} + \mathring{\boldsymbol{X}}_{\tilde{g}}\widehat{\boldsymbol{\beta}}^{\tilde{\boldsymbol{g}}}\right) \tag{8}$$

$$\mathrm{H}_1 : n\widehat{\mathrm{AUC}}_g^{\mathrm{rest}} \neq n\widehat{\mathrm{AUC}}_{\tilde{g}}^{\mathrm{rest}} \iff \frac{1}{T}\ \mathring{\boldsymbol{\omega}}_g^T\left(\mathring{\boldsymbol{X}}_0^{[g]}\widehat{\boldsymbol{\gamma}} + \ \mathring{\boldsymbol{X}}_g\widehat{\boldsymbol{\beta}}^{\boldsymbol{g}}\right) \neq \frac{1}{T}\ \mathring{\boldsymbol{\omega}}\tilde{\boldsymbol{g}}^T\left(\mathring{\boldsymbol{X}}_0^{[\tilde{g}]}\widehat{\boldsymbol{\gamma}} + \mathring{\boldsymbol{X}}_{\tilde{g}}\widehat{\boldsymbol{\beta}}^{\tilde{\boldsymbol{g}}}\right)$$

where $(g, \tilde{g}) \in (1, \cdots, G)^2, g \neq \tilde{g}$ and $\mathring{\boldsymbol{X}}_0^{[g]}$ and $\mathring{\boldsymbol{X}}_g$, respectively, defined as $\boldsymbol{X}_0^{[g]}$ and $\boldsymbol{X}_g$ but restricted to the time interval $[0, T]$. Because $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the parameters of a mixed model and assuming normality hypothesis, it follows that their respective maximum likelihood estimates are approximately normally distributed following the laws $\mathcal{N}(\widehat{\boldsymbol{\beta}}, \widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}))$ and $\mathcal{N}(\widehat{\gamma}, \widehat{\mathrm{Var}}(\widehat{\gamma}))$ and implies that both $\widehat{\boldsymbol{\mu}}_g^{\mathrm{rest}}$ and $n\widehat{\mathrm{AUC}}_g^{\mathrm{rest}}$ are normally distributed. Let note $\widehat{\boldsymbol{\Sigma}}$ the variance-covariance matrix of the estimated fixed parameters given by the inverse of the Fisher information matrix and $\widehat{\boldsymbol{\Sigma}}^{\mathbf{g}}$ the sub-variance covariance matrix of $(\widehat{\boldsymbol{\gamma}}^T, \widehat{\boldsymbol{\beta}}^{\boldsymbol{g}T})^T \in \mathcal{M}_{\dim(\widehat{\gamma})+K_g,1}(\mathbb{R})$. By construction we obtain, $\mathbb{E}(\widehat{\boldsymbol{\mu}}_g^{\mathrm{rest}}) = \ \mathring{\boldsymbol{X}}_0^{[g]}\boldsymbol{\gamma} + \ \mathring{\boldsymbol{X}}_g\boldsymbol{\beta}^{\boldsymbol{g}}$, $\mathrm{Var}(\widehat{\boldsymbol{\mu}}_g^{\mathrm{rest}}) = (\mathring{\boldsymbol{X}}_0^{[g]}\ \mathring{\boldsymbol{X}}_g)\widehat{\boldsymbol{\Sigma}}^{\mathbf{g}}(\mathring{\boldsymbol{X}}_0^{[g]}\ \mathring{\boldsymbol{X}}_g)^T$ and $\mathbb{E}(n\widehat{\mathrm{AUC}}_g^{\mathrm{rest}}) = \frac{1}{T}\mathring{\boldsymbol{\omega}}_g^T\mathbb{E}(\widehat{\boldsymbol{\mu}}_g^{\mathrm{rest}}), \mathrm{Var}(n\widehat{\mathrm{AUC}}_g^{\mathrm{rest}}) = \frac{1}{T^2}\mathring{\boldsymbol{\omega}}_g^T(\mathring{\boldsymbol{X}}_0^{[g]}\ \mathring{\boldsymbol{X}}_g)\widehat{\boldsymbol{\Sigma}}^g(\mathring{\boldsymbol{X}}_0^{[g]}\ \mathring{\boldsymbol{X}}_g)^T\ \mathring{\boldsymbol{\omega}}_g$. Consequently, the asymptotic normal distribution of the estimated difference of the restricted nAUC between the two groups can be inferred with

$$\Delta n\widehat{\mathrm{AUC}}_{g-\tilde{g}}^{\mathrm{rest}} \sim \mathcal{N}\left(\mathbb{E}\left(\Delta n\widehat{\mathrm{AUC}}_{g-\tilde{g}}^{\mathrm{rest}}\right), \mathrm{Var}\left(\Delta n\widehat{\mathrm{AUC}}_{g-\tilde{g}}^{\mathrm{rest}}\right)\right)$$

with $\mathbb{E}(\Delta n\widehat{\mathrm{AUC}}_{g-\tilde{g}}^{\mathrm{rest}}) = \frac{1}{T}\mathring{\boldsymbol{\omega}}_{\tilde{g}T}\mathbb{E}(\widehat{\boldsymbol{\mu}}_{\tilde{g}}^{\mathrm{rest}}) - \frac{1}{T}\ \mathring{\boldsymbol{\omega}}_g^T\mathbb{E}(\widehat{\boldsymbol{\mu}}_g^{\mathrm{rest}})$ and $\mathrm{Var}(\Delta n\widehat{\mathrm{AUC}}_{g-\tilde{g}}^{\mathrm{rest}}) = \ \mathring{\boldsymbol{\omega}}^{\mathrm{T}}(\mathring{\boldsymbol{X}}_0\ \mathring{\boldsymbol{X}})\widehat{\boldsymbol{\Sigma}}(\mathring{\boldsymbol{X}}_0\ \mathring{\boldsymbol{X}})\ \mathring{\boldsymbol{\omega}}$, $\mathring{\boldsymbol{\omega}} \in \mathcal{M}_{\mathring{m}_g+\mathring{m}\tilde{g},1}(\mathbb{R})$ being defined as $\frac{1}{T}(\boldsymbol{0}^{\mathrm{T}}, \mathring{\boldsymbol{\omega}}_{\tilde{g}}^T)^T - \frac{1}{T}(\mathring{\boldsymbol{\omega}}_g^T, \boldsymbol{0}^{\mathrm{T}})^T$. For a test of the null hypothesis defined in equation (8), we can build the standard normally distributed Z-statistic given by

$$Z = \frac{\Delta n\widehat{\mathrm{AUC}}_{g-\tilde{g}}^{\mathrm{rest}}}{\sqrt{\mathrm{Var}\left(\Delta n\widehat{\mathrm{AUC}}_{g-\tilde{g}}^{\mathrm{rest}}\right)}}$$

Under the null hypothesis, the Z-statistics follows a $\mathcal{N}(0, 1)$. By weighted averaging incomplete measures, the impact of potential heteroscedasticity is reduced due to the AUC-based approach. If still variance heterogeneity

between the group occur, the Z-statistics can be modified into a Student's t-test like statistics with degree of freedom τ (equals to ∞ in case of Z-statistic). As matter of fact, in case of remaining heterogeneity, data specific to each group should be fitted with specific and independent mixed effects model. The T-statistic resulting from this procedure will differ from our Z-statistic by its standard deviation simply defined as the squared root of the sum of the variances of the group-specific nAUC, and with a degree of freedom defined by the Satterthwaite approximation[41,42]

$$\tau = \frac{\left(\mathrm{Var}\left(n\widehat{\mathrm{AUC}}_g^{\mathrm{rest}}\right) + \mathrm{Var}\left(n\widehat{\mathrm{AUC}}_{\bar{g}}^{\mathrm{rest}}\right)\right)^2}{\frac{\mathrm{Var}\left(n\widehat{\mathrm{AUC}}_g^{\mathrm{rest}}\right)}{n_g - 1} + \frac{\mathrm{Var}\left(n\widehat{\mathrm{AUC}}_{\bar{g}}^{\mathrm{rest}}\right)}{n_g - 1}}$$

Similarly, in case of small sample size, our Z-test can be modified into Student's t-test with degree of freedom defined by the Kenward-Roger approximation.[43] Similarly to Bailer,[44] a $100(1-\alpha)\%$ confidence interval for $\Delta n\widehat{\mathrm{AUC}}_{g-\bar{g}}^{\mathrm{rest}}$ can be derived from the statistic, as

$$\Delta n\widehat{\mathrm{AUC}}_{g-\bar{g}}^{\mathrm{rest}} \pm z_{\tau,\alpha/2}\sqrt{\mathrm{Var}\left(\Delta n\widehat{\mathrm{AUC}}_{g-\bar{g}}^{\mathrm{rest}}\right)}$$

where $z_{\tau,\alpha/2}$ is the $(1-\alpha/2)100^{th}$ percentile of the distribution.

An extension to k-sample design is straightforward deriving a one-way ANOVA testing the equality of normalized AUCs. Similarly to our Z-statistics, nAUCs are compared on the same interval of calculation $[0, T]$ with $T = \min_{g\in\{1,\cdots,G\}}(T_g)$.

$$\begin{cases} H_0: & n\widehat{\mathrm{AUC}}_1^{\mathrm{rest}} = n\widehat{\mathrm{AUC}}_2^{\mathrm{rest}} = \cdots = n\widehat{\mathrm{AUC}}_K^{\mathrm{rest}}, \\ H_1: & \exists(i,j)\,|\, n\widehat{\mathrm{AUC}}_i^{\mathrm{rest}} \neq n\widehat{\mathrm{AUC}}_j^{\mathrm{rest}} \end{cases}$$

where K is the number of groups compared by the k-sample test, $K \leq G$. Similarly to classic one-way ANOVA, we define the statistic F following Fisher law as

$$F = \frac{\frac{SS_{between}}{K-1}}{\frac{SS_{within}}{N_K - K}} \sim F(K-1, N_K - K)$$

where $N_K = \sum_{g=1}^{K} n_g$ and $SS_{between}$ and SS_{within} define respectively the inter- and intra-group variability and are calculated as

$$SS_{between} = \sum_{g=1}^{K} n_g \left(n\widehat{\mathrm{AUC}}_g^{\mathrm{rest}} - \frac{1}{K}\sum_{k=1}^{K} n\widehat{\mathrm{AUC}}_k^{\mathrm{rest}}\right)^2$$

$$SS_{within} = \sum_{g=1}^{K} n_g^2 \mathrm{Var}(n\widehat{\mathrm{AUC}}_g^{\mathrm{rest}})$$

4 Simulation study

In this section, we conduct a simulation study to analyze the statistical properties of our approach. The simulation setting is driven by the motivating examples described in section 2.

4.1 Generation of simulated data

We simulate longitudinal data mimicking a randomized HIV therapeutic vaccine trial involving two groups of treatment in which the outcome of interest is the HIV RNA load measurement. We simulated data using a LMEM as described by (9)

$$\begin{aligned} Y_{ij,g} = \gamma_0 + \mathbb{1}_{[g=1]} \sum_{k=1}^{K_1} \beta_k^1 \phi_k^1(t_{ij,1}) + \mathbb{1}_{[g=2]} \sum_{k=1}^{K_2} \beta_k^2 \phi_k^2(t_{ij,2}) \\ + b_{0i} + \sum_{k=1}^{K_i} b_{ki} \Psi_k^i(t_{ij,g}) + \varepsilon_{ij} \end{aligned} \tag{9}$$

where $Y_{ij,g}$ is the outcome of the ith subject belonging to the gth group at the jth time point where $i \in \{1, \cdots, n_g\}$, $j \in \{1, \cdots, m_g\}$ and $g \in \{1, 2\}$. In this model, the non-group-specific function f_0 is a global intercept labeled γ_0, while random effects are described by individual smooth cubic B-splines curves defined as linear combination of the cubic B-spline basis $\boldsymbol{\Psi^i} = (\Psi_1^i, \cdots, \Psi_{K_i}^i)^T$ with $\boldsymbol{b_i} = (b_{1i}, \cdots, b_{K_i i})^T$ as regression coefficients, $\forall i \in \{1, \cdots, N\}$, $N = n_1 + n_2$. Similarly, the group-specific fixed effects are modeled by cubic B-spline curves with $\boldsymbol{\phi^g} = (\phi_1^g, \cdots, \phi_{K_g}^g)^T$ and $\boldsymbol{\beta^g} = (\beta_1^g, \cdots, \beta_{K_g}^g)^T$ as spline basis and regression coefficients, respectively. Random effects describing the inter-individual variability are assumed to be normally distributed $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0}, \Omega)$ as well as the error terms $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_e^2)$. Based on the HIV RNA load data from the Vac-IL2 trial (see section 2, *Motivating Examples*), we evaluated the regression coefficient estimates $\gamma_0, \boldsymbol{\beta^1}, \boldsymbol{\beta^2}$ and $\boldsymbol{b}$ as well as the parameters K_g and K_i being respectively the number of spline basis involved in the group-specific and individual spline curves. The model involving a global intercept γ_0, the splines basis have been built without including intercept terms making K_g and K_i equal to the sum of the number of internal knots and the degree (fixed at 3 in our case) of the respective spline curves.

For the purpose of examining the properties of the proposed approach developed to test the equality of nAUCs, we generate numerous vaccine trials. As illustrated in Figure 1, we simulated two types of mean trajectory profiles: one in which the timing of viral rebound is similar in control and treatment group but the magnitude of the rebound may differ, and one in which the timing of viral rebound is expected to be longer in the treatment group compared to the control group. Finally, outcomes are measured at a constant time interval such as $t = (0, 1, 2, \cdots, 24)^T$ weeks and the number of patients by group $n = n_1 = n_2$ varied amongst 20, 50 and 100. They reproduce the trajectories found in the Vac-IL2 and LIGHT trials (see section 2, *Motivating Examples*). Based on the Vac-IL2 data, we set the values of $\sigma_e^2 = 0.2$, the fixed intercept $\gamma_0 = -0.44$ and the fixed parameters of the first group of treatment ($g = 1$) seen as the control group, $\boldsymbol{\beta^1}$ (see Table 1). The five fixed parameters of the treatment group in both profiles $\boldsymbol{\beta^2}$ have been chosen such as given values of $\Delta n\text{AUC}_{1-2}$ are targeted to specific values. To test the properties of the method, we simulated data with $\Delta n\text{AUC}_{1-2}$ taking values of 0, –0.1 and –0.25 $\log_{10}$ cp/ml. We defined the number of fixed splines basis as $K_1 = K_2 = 5$ for both profiles with the two internal knots fixed at $(0.25, 5.62)$ weeks for both groups in profile 1 and $(0.25, 5.62)$ and $(3.23, 7.63)$ weeks in profile 2 for control and vaccine groups, respectively. Similarly, we fixed the number of random spline basis $K_i = 5$ with $(2.0, 4.5)$ weeks as internal knots in profile 1 and $(2.0, 4.5)$ and $(5.0, 8.0)$ weeks in Profile 2 for control and vaccine groups, respectively. Number and positions of internal knots have been optimally chosen on Vac-IL2 data by applying the R-package *freeknotspline*[45] using AIC as optimization criterion.

The covariance matrix of the random effects Ω is defined as diagonal such as $\Omega = \sigma_b^2 \mathbf{I}_{K_i+1}$ where the value of σ_b^2 has been chosen according to the targeted values of $\text{Var}(n\text{AUC}_g)$. The estimated variances of nAUC were 0.027 and 0.021, respectively, in the control and the treatment group in Vac-IL2 trial. Hence, in simulations, we tested the impact of the intra-group variability when $\text{Var}(n\text{AUC}_g)$ was equal to 0.02 and 0.1, in both groups.

We generated MAR monotonic missing data as follows. For each subject i at each time point j, the outcome $Y_{ij,g}$ was labeled as missing if $Y_{ij,g} \in \{Y_{ij,g} | \exists j' \leq j, \{Y_{ij',g} \geq \alpha\} \cap \{Y_{ij'-1,g} \geq \alpha\}\}$, with α being a fixed threshold. A patient dropped out from the trial if his/her HIV RNA load exceeded the threshold α at two consecutive time points. The subsequent measurements were considered as missing. We investigated the impact of the missing data on the robustness of the method by considering three values for the threshold α: 100,000 ($5\log_{10}$), 50,000 ($\sim 4.7\log_{10}$) and 10,000 ($4\log_{10}$) cp/ml. As illustrated in Figure 2 for the profile 1, the percentage of drop-out in each trial was inversely linked to the value of α. Due to the difference of nAUC between the two groups, each

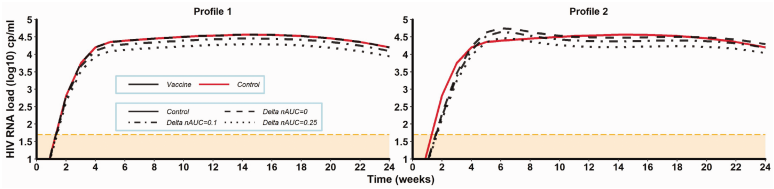


Figure 1. Simulated mean trajectories of HIV RNA load over time for both profiles 1 and 2. Note: Red solid line represents Group 1 (Control), dashed, dot dashed and dotted lines represent Group 2 (treatment) when ΔnAUC with Group 1 is equal to 0, –0.1 and –0.25, respectively. Orange dashed line and area delimit the LOD $= \log_{10}(50)$. LOD: limit of detection.

Table 1. Fixed parameter values used to simulate control and vaccine groups for both profiles, according to ΔnAUC values.

Treatment group	Profile 1	Profile 2
Control group, β^1	$(-0.55, 4.72, 4.96, 5.18, 4.64)$	$(-0.55, 4.72, 4.96, 5.18, 4.64)$
ΔnAUC$= 0$, β^2	$(-0.55, 4.72, 4.96, 5.18, 4.64)$	$(1.38, 5.57, 4.53, 5.20, 4.74)$
ΔnAUC$= 0.1$, β^2	$(-0.54, 4.61, 4.85, 5.07, 4.54)$	$(1.35, 5.44, 4.43, 5.09, 4.63)$
ΔnAUC$= 0.25$, β^2	$(-0.52, 4.46, 4.69, 4.90, 4.39)$	$(1.31, 5.26, 4.28, 4.92, 4.48)$

Note: The value of the global intercept was fixed at $\gamma_0 = -0.44$.

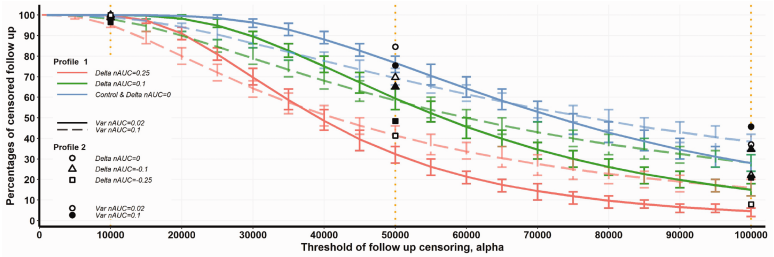


Figure 2. Percentages of censored follow-up when data simulated by both profiles are impacted by the threshold of lost of follow-up α. Note: Lines display percentages obtained for the profile 1 with solid and dashed lines representing data simulated with Var(nAUC) $= 0.02$ and 0.1, respectively. Blue lines describe both Group 1 (Control) and Group 2 (treatment) when ΔnAUC with Group 1 is equal to 0, green and pink lines represent Group 2 when ΔnAUC $= 0.1$ and 0.25, respectively. Marks display percentages obtained for the Profile 2 with empty and full marks representing data simulated with Var(nAUC) $= 0.02$ and 0.1, respectively. The squares, triangles and circles describe Group 2 when ΔnAUC $= 0$, 0.1 and 0.25 with the control group in blue, respectively. Vertical dotted lines highlight the positions of $\alpha = 100{,}000$, 50,000 and 10,000 cp/ml.

value of α generated both equal (ΔnAUC $= 0$, blue curves) and unequal (ΔnAUC $\neq 0$, blue curve for control and green/pink curves for treatment group) drop-out rates. While $\alpha = 100{,}000$ cp/ml leaded to approximately 30% of drop-out in control group and respectively 30%, 15% and 5% in treatment group when $\Delta n\text{AUC} = 0, 0.1$ and 0.25, for Var(nAUC) $= 0.02$, these percentages increased respectively until 75%, 75%, 60% and 35% for $\alpha =$ 50,000. Finally, the choice of $\alpha = 10{,}000$ allowed to test the method with extremely high percentages of drop-out which were in the neighborhood of 100%. The consideration of the second profile of data simulation leaded to a slight increase of these percentages of approximately 7% when the variance of nAUC was equal to 0.1 and 10% for 0.02.

We also generated left-censored outcomes using the limit of detection for viral load at $50 \sim 1.7\log_{10}$ cp/ml, which has been chosen in accordance with values typically encountered in our motivating examples. This choice of LOD generated mean percentages of undetectable data in each group ranging from 7.30% to 7.70% for profile 1 and from 7.30% to 8.70% for profile 2, representing approximately two time points with undetectable outcome over 25.

4.2 Analysis of simulated data

We analyzed the simulated data using a well-specified model. Formulas for nAUC are derived from equation (9). MEM estimations took into account left-censored outcomes using an hybrid EM-algorithm implemented in the R-package *lmec*.[46] Let note $(\widehat{\gamma}_0, \widehat{\boldsymbol{\beta}}^1, \widehat{\boldsymbol{\beta}}^2)^T$ the vector of the estimated fixed parameters where $\widehat{\boldsymbol{\beta}}^g = (\widehat{\beta}^g_1, \cdots, \widehat{\beta}^g_{K_g})^T$, for $g \in \{1, 2\}$. Using the model in equation (9), the expected value of Y in the gth group at any time $t_{j,g}$ is $\widehat{\mu}_{j,g} = \widehat{\gamma}_0 + \sum_{k=1}^{K_g} \widehat{\beta}^g_k \phi^g_k(t_{j,g})$, which allows to approximate the nAUC in each group, its variance and the difference in nAUC as follows

$$n\widehat{\mathrm{AUC}}_g = K_{\gamma g}\widehat{\gamma}_0 + \sum_{k=1}^{K_g} \widehat{\beta}^g_k C_{kg}$$

$$\Delta n\widehat{\mathrm{AUC}}_{1-2} = \widehat{\gamma}_0(K_{\gamma 2} - K_{\gamma 1}) + \sum_{k=1}^{K_2} \widehat{\beta}^2_k C_{k2} - \sum_{k=1}^{K_1} \widehat{\beta}^1_k C_{k1}$$

$$\mathrm{Var}(n\widehat{\mathrm{AUC}}_g) = (K_{\gamma g})^2 \mathrm{Var}(\widehat{\gamma}_0) + \sum_{k=1}^{K_g} (C_{kg})^2 \mathrm{Var}(\widehat{\beta}^g_k) + 2\sum_{k=1}^{K_g - 1} \sum_{\tilde{k}=k+1}^{K_g} C_{kg} C_{\tilde{k}g} \mathrm{Cov}(\widehat{\beta}^g_k, \widehat{\beta}^g_{\tilde{k}})$$

$$+ 2\sum_{k=1}^{K_g} K_{\gamma g} C_{kg} \mathrm{Cov}(\widehat{\gamma}_0, \widehat{\beta}^g_k)$$

where C_{kg} and $K_{\gamma g}$ are defined by $C_{kg} = \frac{1}{T_g} \sum_{j=2}^{m_g} \frac{(t_{j,g} - t_{j-1,g})}{2} [\phi^g_k(t_{j,g}) + \phi^g_k(t_{j-1,g})]$ and $K_{\gamma g} = \frac{2}{T_g} \sum_{j=2}^{m_g} \frac{(t_{j,g} - t_{j-1,g})}{2}$.

For each combination of simulated datasets and missing data patterns, 1000 replications were performed with the objective of evaluating the robustness of the method to test the equality of areas under the curves between the two groups through its type-I error, its power and the bias in the estimation of the difference of nAUC. We compared the results provided by our method with a standard two-sample t-test for the difference of nAUC between the two groups, i.e. $H_0 : \overline{n\mathrm{AUC}_2} - \overline{n\mathrm{AUC}_1} = 0$ where $\overline{n\mathrm{AUC}_g} = \frac{1}{n_g} \sum_{i=1}^{n_g} n\mathrm{AUC}_i$ with $n\mathrm{AUC}_i$ defined by equation (1). We performed this test without accounting for missing data and using two common ad hoc approaches: the last observation carried forward (LOCF) where missing data are imputed by the last observed value before the follow-up censoring, and the mean imputation where missing observations are imputed by the mean of the observations before this follow-up censoring.

In addition to the standard two-sample t-test, we compared our method with the t-test version of the non-parametric two-sample test proposed by Vardi et al.[25] This test was developed to compare a one-dimensional variable such as AUC between two groups of treatment when individual follow-up is subject to informative homogeneous or heterogeneous censoring. In order to be able to compare the results provided by this test and our method, we applied this test to normalized AUC. The test is based on U-statistics defined as

$$U_{m_1,m_2} = \frac{1}{m_1 m_2} \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} D_{i_1,i_2}$$

where m_1 and m_2 are respectively the number of subjects in the first and the second compared groups, g_1 and g_2, while D_{i_1,i_2} is defined as the paired cross-treatment contrast for the cross-treatment pair $(i_1, i_2) \in g_1 \times g_2$

$$D_{i_1,i_2} = \frac{1}{T_{i_1} \wedge T_{i_2}} \int_0^{T_{i_1} \wedge T_{i_2}} \left[Y_{i_2,g_2}(t) - Y_{i_1,g_1}(t) \right] \mathrm{d}t$$

$$= \frac{1}{T_{i_1} \wedge T_{i_2}} \left[\mathrm{AUC}_{i_2}\Big|_{[0, T_{i_1} \wedge T_{i_2}]} - \mathrm{AUC}_{i_1}\Big|_{[0, T_{i_1} \wedge T_{i_2}]} \right]$$

where $T_{i_1} \wedge T_{i_2} = \min(T_{i_1}, T_{i_2})$. The variable D_{i_1,i_2} is then defined as the difference of nAUC between the subjects i_1 and i_2, restricted to their common time of follow-up. Similarly to the simulation studies conducted in their paper, we defined the variance of the U-statistic as equation (2.15) in Vardi's paper[25]

$$\widehat{\sigma}^2_{m_1,m_2} = \sum_{i_1=1}^{m_1} \frac{\left(\overline{D}_{i_1\cdot} - \overline{D}_{\cdot\cdot}\right)^2}{m_1(m_1-1)} + \sum_{i_2=1}^{m_2} \frac{\left(\overline{D}_{\cdot i_2} - \overline{D}_{\cdot\cdot}\right)^2}{m_2(m_2-1)}$$

where $\overline{D}_{i_1\cdot} = \sum_{i_2} D_{i_1,i_2}/m_2$, $\overline{D}_{\cdot i_2} = \sum_{i_1} D_{i_1,i_2}/m_1$ and $\overline{D}_{\cdot\cdot} = U_{m_1,m_2}$ and we considered the following null hypothesis H_0 : *the distribution of D is symmetric about 0*.

Five procedures are then compared for testing the equality of nAUC including three ad hoc methods respectively called Indiv. nAUC Data, Indiv. nAUC LOCF and Indiv. nAUC Mean Imp., the non-parametric test called NP nAUC and our approach called MEM nAUC.

4.3 Simulation results

The results of our simulations evaluating the robustness of the test of equality of nAUC are displayed in Table 2. Although only results for simulations involving $n_g = 50$ patients by group are presented in the main body of the article, extended results for $n_g = 20$ and 100 can be found in Online Appendix C, Tables C.2a and C.3a. In these simulations, as expected with a well-specified model, when there is no censored follow-up and no left censoring using individual nAUC, non-parametric approach or our method based on MEM nAUC are identical in term of type-I error, which are kept to their nominal level of 5% (between 0.044 and 0.06). However, the power seems to be consistently higher for MEM nAUC in particular when the inter-individual variability is high. When introducing the LOD at 50 cp/ml, the results are similar for profile 1 but tend to show a superiority of MEM nAUC for profile 2 in which there are a larger number of left-censored observations due to delay in viral rebound in one group. This is explained by the fact that MEM nAUC, contrary to individual nAUC involved either in indiv. nAUC or NP nAUC methods, accounts for left censoring instead of considering censored data fixed to their censorship level value. When the threshold of HIV RNA defining drop-out, α, is equal to 100,000 and 50,000 cp/ml, all individual methods (with or without adjustment for missing data) fail in term of type-I error in the second profile with lagged increasing trajectories of viral load (see Figure 1). Even when the type-I error is controlled such as for profile 1 (with the same shape of mean trajectories see Figure 1), the power for raw data and mean imputation approaches is low for most settings. While the NP nAUC method shows controlled type-I error between 0.048 and 0.057 for profiles 1 and 2 when α is equal to 100,000 cp/ml and for profile 1 when the threshold is equal to 50,000 cp/ml, we observe an inflation of the type-I error up to 0.075 for the second profile. On the contrary, the MEM nAUC method shows type-I error between 0.048 and 0.064 for profiles 1 and 2. When variability is low, the power is also good and higher than 76% for the two methods. In all cases, the power found in these settings is similar in magnitude to the power obtained when there is no censored follow-up and no left censoring for viral load. When the threshold α is equal to 10,000 copies/ml, while all individual methods and the non-parametric approach fail to control the type-I error for the profile 2, our approach MEM nAUC successfully gets a type-I error around the nominal value for both profiles. This result is mainly driven by the difference of the shapes of the mean trajectories for the two compared groups in Profile 2. In fact, as shown in Figure 1, the difference of nAUC appears as quite homogeneously distributed over the time of follow-up in profile 1 leading to robust results for all methods despite an early drop out for a high percentage of subjects. However, in profile 2, the value of ΔnAUC resulting from the compensation of the beginning and the end of the dynamics, only the parametric method is able to capture the true difference of nAUC regardless of the premature censored follow-up for more than 80% of individuals.

In addition, we graphically illustrated the estimated bias and standard error for ΔnAUC obtained for each method in Figure 3. For all profiles, when there is no drop-out or when the threshold α is high enough (equal to 100,000 and 50,000 cp/ml), the bias is closer to 0 for MEM nAUC compared to other methods. Also, the standard error of ΔnAUC calculated with MEM nAUC is similar to the non-parametric approach and closer to all the ad hoc individual methods to the theoretical values of standard error of ΔnAUC, respectively 0.028 for $\mathrm{Var}(n\mathrm{AUC}) = 0.02$ and 0.063 for $\mathrm{Var}(n\mathrm{AUC}) = 0.1$. This mostly explains the comparable robustness between MEM nAUC and NP nAUC and their better performances in term of power compared to individual methods. When α is equal to 10,000 cp/ml, the inflated type-I errors observed for individual and non-parametric methods

Table 2. Comparison of the robustness of the test of equality of nAUC calculated as individual summary measures and mixed model summary statistics.

			Profile 1						Profile 2					
		Methods	Type-I error		Power				Type-I error		Power			
Data pattern		ΔnAUC	0		−0.1		−0.25	0		−0.1		−0.25		
LOD	α	Var(nAUC)	0.02	0.1	0.02	0.1	0.02	0.1	0.02	0.1	0.02	0.1	0.02	0.1
		Indiv. nAUC	0.060	0.060	0.95	0.33	1.00	0.96	0.046	0.053	0.94	0.35	1.00	0.97
		NP nAUC	0.060	0.046	0.95	0.33	1.00	0.96	0.046	0.053	0.94	0.35	1.00	0.97
∅	∅	MEM nAUC	0.059	0.055	0.96	0.41	1.00	0.99	0.044	0.056	0.95	0.44	1.00	1.00
		Control	0	0	0	0	0	0	0	0	0	0	0	0
Mean missing rate (%)[a]		Treatment	0	0	0	0	0	0	0	0	0	0	0	0
		Indiv. nAUC	0.056	0.049	0.96	0.35	1.00	0.97	0.062	0.063	0.89	0.30	1.00	0.97
		NP nAUC	0.056	0.049	0.96	0.35	1.00	0.97	0.062	0.063	0.89	0.30	1.00	0.97
50	∅	MEM nAUC	0.063	0.053	0.95	0.35	1.00	0.97	0.047	0.055	0.94	0.37	1.00	0.97
		Control	0	0	0	0	0	0	0	0	0	0	0	0
Mean missing rate (%)		Treatment	0	0	0	0	0	0	0	0	0	0	0	0
		Indiv. nAUC												
		1. Data	0.060	0.054	0.49	0.16	1.00	0.79	0.540	0.526	0.92	0.37	1.00	0.97
		2. LOCF	0.052	0.045	0.84	0.32	1.00	0.96	0.281	0.170	0.31	0.11	1.00	0.84
		3. Mean Imp.	0.059	0.053	0.51	0.16	1.00	0.80	0.529	0.500	0.83	0.65	1.00	0.94
		NP nAUC	0.053	0.053	0.93	0.32	1.00	0.96	0.057	0.053	0.79	0.32	1.00	0.96
	1.10^5	MEM nAUC	0.064	0.054	0.94	0.33	1.00	0.96	0.053	0.060	0.92	0.35	1.00	0.97
		Control	28	38	28	38	27	39	28	38	28	38	28	39
Mean missing rate (%)		Treatment	28	38	15	28	5	16	37	46	22	35	8	21
		Indiv. nAUC												
		1. Data	0.050	0.052	0.05	0.05	0.13	0.16	0.946	0.881	0.82	0.79	0.70	0.78
		2. LOCF	0.046	0.051	0.77	0.29	1.00	0.95	0.483	0.233	0.11	0.06	1.00	0.71
		3. Mean Imp.	0.051	0.050	0.05	0.05	0.14	0.17	0.940	0.845	0.81	0.76	0.70	0.77
		NP nAUC	0.048	0.053	0.81	0.27	1.00	0.93	0.069	0.075	0.77	0.44	1.00	0.97
	5.10^4	MEM nAUC	0.063	0.060	0.84	0.31	1.00	0.95	0.048	0.051	0.76	0.31	1.00	0.96
		Control	77	69	77	69	77	69	77	69	77	69	77	69
Mean missing rate (%)		Treatment	77	69	59	58	32	41	85	76	70	65	41	48
		Indiv. nAUC												
		1. Data	0.041	0.057	0.04	0.06	0.12	0.07	0.894	0.868	0.91	0.87	0.85	0.80
		2. LOCF	0.058	0.043	0.20	0.15	0.81	0.68	0.555	0.421	0.18	0.11	0.13	0.17
		3. Mean Imp.	0.039	0.050	0.04	0.06	0.10	0.07	0.746	0.725	0.83	0.76	0.80	0.73
		NP nAUC	0.055	0.053	0.43	0.19	1.00	0.77	0.972	0.651	1.00	0.92	1.00	1.00
	1.10^4	MEM nAUC	0.059	0.058	0.31	0.19	0.91	0.60	0.051	0.073	0.23	0.19	0.84	0.58
		Control	100	99	100	99	100	99	100	99	100	99	100	99
Mean missing rate (%)		Treatment	100	99	100	98	100	95	100	100	100	99	100	97

AUC: area under the curve; nAUC: normalized AUC; NP: non parametric; Individual ad hoc methods (Indiv. nAUC): 1. Data: raw data, 2. LOCF: last observation carried forward, 3. Mean Imp.: mean imputation.

[a] Missing rate: Percentage of subjects dropping out before the end of the study.

Note: Individual trajectories are subject to missing data and/or limit of detection. Simulations were performed for $n_g = 50$ subjects by group, mean trajectories following both profiles and for 1000 replications.

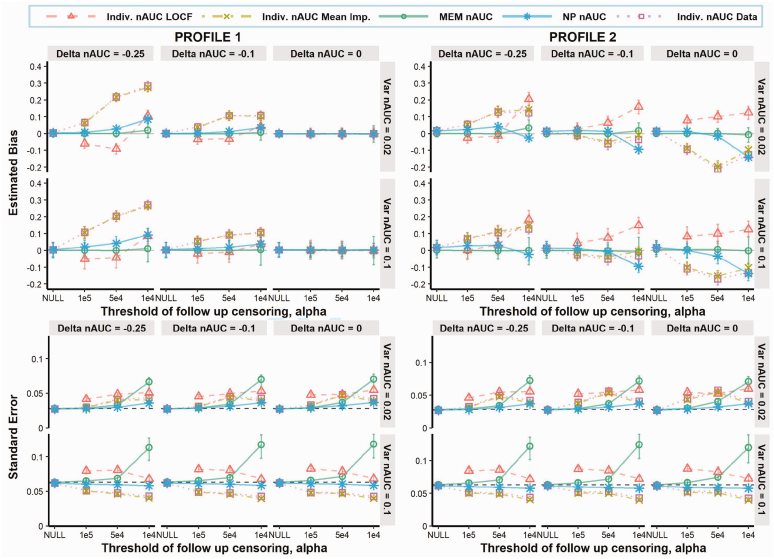


Figure 3. Comparison of the estimated bias and standard error of ΔnAUC obtained by the three individual methods Indiv. nAUC Data, Indiv. nAUC LOCF, Indiv. nAUC Mean Imp., the non-parametric test Non Param. and our method MEM nAUC. Both criteria were estimated for data subject to a LOD, with or without censored follow-up, with $n_g = 50$ subjects by group, mean trajectories following both profiles, for 1000 replications. Note: Pink dashed lines and triangles represent Ind. nAUC LOCF, green dot-dashed lines and crosses represent Indiv. nAUC Mean Imp., green solid lines and circles represent MEM nAUC, blue solid lines and stars represent NP nAUC and purple dotted lines with squares represent Indiv. nAUC Data. In standard Error plots, black dashed lines display the theoretical values (0.028 when Var(nAUC) = 0.02 and 0.063 for 0.1); LOCF: last observation carried forward.

are explained by biased estimates of ΔnAUC which are not compensated by an increased value of the standard error, unlike the MEM nAUC method.

4.4 Relaxing the correct model specification assumption

The validity of the method relies on the correct specification of the MEM as described in equation (3) in the section Method. To relax this assumption, we conducted additional simulations to evaluate the method when data are fitted with another MEM. To evaluate the performances in a setting closer to real-data, the number and position of the knots in the MEM defined in equation (9) were also estimated with the data. We used the R-package *freeknotspline* to estimate and replace the two sets of fixed two internal knots (2.0, 4.5) and (5.0, 8.0) involved in the build of group-specific spline curves by a set of knots optimizing the fit of data. Moreover, spline basis was built with external knots chosen as (0, T_g) instead of (0, 24) considering the real observed time of follow-up, which can be modified with censored follow-up. For each simulation, the number of internal knots for a given group is optimized between 1 and 3 as well as their position using AIC as optimization criterion. Three other selection criteria have been tested: BIC, adjAIC, adjGCV and compared to AIC. Similar results of power and type-I error have been obtained for the four criteria (results not shown). Spline basis involved in random effects were similarly built chosen (0, T_i) as boundary knots and the number of internal knots chosen between 1 and 2. This adaptive feature of the model allows to build group-specific spline basis taken into account both left-censored and missing data. The results obtained by this model are displayed in Table 3 for $n_g = 50$ subjects by group. Similar results are presented in Online Appendix C in Tables C.2b and C.3b for $n_g = 20$ and 100, respectively.

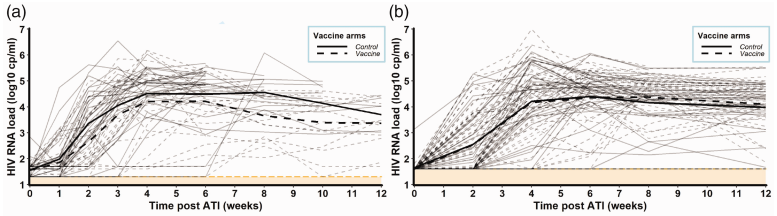
In all settings except for high level of censored follow-up with $\alpha = 10{,}000$, using adaptive MEM led to equivalent type-I error (between 0.046 and 0.063 instead of 0.044 and 0.064) and power than with the well-specified model, for both profiles. Using adaptive MEM slightly increased the type-I error when the threshold for drop-out

Table 3. Robustness of the test of equality of nAUC calculated as mixed model summary statistics considering the MEM (9) with adaptive spline basis.

			Profile 1						Profile 2					
			Type-I error		Power				Type-I error		Power			
Data pattern		Method ΔnAUC	0		−0.1		−0.25	0		−0.1		−0.25		
LOD	α	Var(nAUC)	0.02	0.1	0.02	0.1	0.02	0.1	0.02	0.1	0.02	0.1	0.02	0.1
∅	∅	Adap. MEM	0.060	0.060	0.96	0.41	1.00	0.99	0.046	0.054	0.95	0.42	1.00	1.00
50	∅	Adap. MEM	0.063	0.056	0.95	0.35	1.00	0.97	0.050	0.051	0.94	0.37	1.00	0.98
	1.10^5	Adap. MEM	0.060	0.054	0.94	0.33	1.00	0.96	0.050	0.060	0.91	0.35	1.00	0.97
	5.10^4	Adap. MEM	0.060	0.059	0.84	0.31	1.00	0.95	0.049	0.050	0.77	0.30	1.00	0.95
	1.10^4	Adap. MEM	0.070	0.061	0.31	0.17	0.89	0.54	0.061	0.078	0.26	0.17	0.83	0.54

AUC: area under the curve; nAUC: normalized AUC; LOD: limit of detection.

Note: Individual trajectories are subject to missing data and/or LOD. Simulations were performed for $n_g = 50$ subjects by group, mean trajectories following both profiles and for 1000 replications.



	Number of LOD by patient mean (IQR)	% LOD
Vac-IL2 Trial		
Control	1.903 (1.00-2.00)	28.2%
Vaccine	2.406 (1.75-3.00)	33.5%
LIGHT Trial		
Control	1.656 (1.00-2.00)	27.9%
Vaccine	1.763 (1.00-2.00)	29.8%

Figure 4. Exploratory plots and table for the control and vaccine groups from the Vac-IL2 and LIGHT HIV therapeutic vaccine trials. Observations are subject to LODs of 40 cp/ml or 20 and 50 cp/ml for LIGHT and Vac-IL2 trial, respectively. LOD: limit of detection. (a) Outcome trajectories for the control and vaccine groups of the Vac-IL2 HIV therapeutic vaccine trial, with two LOD $= \log_{10}(50)$ and $\log_{10}(20)$ cp/ml. (b) Outcome trajectories for the control and vaccine groups of the LIGHT HIV therapeutic vaccine trial, with LOD $= \log_{10}(40)$ cp/ml. (c) Mean number by patient and global percentage of observations below the LOD. (d) Percentage of missing data over time. Note: In (a) and (b), thick lines describe mean dynamics and thin lines individual ones, solid lines represent control group and dashed lines represent vaccine group. In (d), black lines with circles describe data from LIGHT trial, grey lines with crosses describe data from Vac-IL2 trial, solid lines represent control groups and dashed lines represent vaccine groups.

is 10,000 (between 0.061 and 0.078 instead of 0.051 and 0.073), while the estimated power remained unchanged. Altogether, even when the MEM structure is not known, this simulation shows that it is possible to use adaptive MEM for the modeling of the marker trajectories without invalidating the method, making it more relevant on real data.

5 Application on real clinical data

As illustrative examples, we applied the presented approach to the log-transformed HIV RNA load data from the Vac-IL2 and LIGHT trials (see section 2, Motivating Examples). Exploratory plots of the individual and mean HIV RNA load dynamics for control and vaccine groups are shown in Figure 4(a) and (b), for VAC-IL2 and LIGHT trials, respectively. As illustrated in table in Figure 4(c), longitudinal data in both trials are subject to left-censoring. While two values of LOD are considered in Vac-IL2 trial, 20 and 50 cp/ml (~ 1.3 and $1.7\log_{10}$ cp/ml), impacting a total of 28.2% and 33.5% of observations for control and vaccine groups, only a LOD at 40 cp/ml ($\sim 1.6\log_{10}$ cp/ml) is involved in LIGHT trial, leading to 27.9% and 29.8% of observations in the respective groups. In addition to left-censoring, those data are impacted by drop-outs. In LIGHT trial, ART resumption was required in case of serious AIDS or non-AIDS adverse events, when two consecutive of CD4+ T cells counted below 350 cells/mm^3 within at least a two weeks' time interval as well as for specific patient or physician willingness. Approximately 20% of patients were concerned by these rules and resumed ART before the end of the predefined 12 weeks of ATI (see Figure 4(d)) being considered as drop-outs. In Vac-IL2 trial, 63% and 84% of drop-outs occurred in vaccine and control group respectively, as the result of HIV RNA load exceeding 50,000 cp/ml at four or six weeks post-ATI or exceeding 10,000 cp/ml after eight weeks of ART interruption.

We applied the proposed approach discussed in the manuscript using the MEM described by equation (9) where the number and the position of internal knots for both population and individual levels are optimized on data using the R-package *freeknotspline* and AIC criteria. Also, the structure of the covariance matrix of random effects being unknown, we estimated this matrix as unstructured instead of diagonal. Moreover, we verified the applicability of our method on these real data by checking the normality of the distribution of the residuals provided by the MEM as well as the homoscedasticity of its error model for both trials (see Online Appendix E). We compared the results obtained by our approach, where the difference of nAUC between the two groups of

Table 4. Summary of results from both Vac-IL2 and LIGHT studies.

	Estimate (SE)	95% CI	p-value	Estimate (SE)	95% CI	p-value
Methods	Vac-IL2 trial			LIGHT trial		
Data	−0.346 (0.170)	[−0.680; −0.013]	0.046	−0.030 (0.175)	[−0.312; 0.372]	0.864
LOCF	−0.382 (0.198)	[−0.770; 0.007]	0.060	−0.018 (0.186)	[−0.382; 0.346]	0.924
Mean Imp.	−0.345 (0.312)	[−0.957; 0.266]	0.276	0.217 (0.245)	[−0.263; 0.697]	0.959
NP nAUC	−0.349 (0.205)	[−0.751; 0.053]	0.089	0.042 (0.178)	[−0.306; 0.390]	0.813
Adap. MEM	−0.459 (0.213)	[−0.877; −0.041]	0.031	0.095 (0.216)	[−0.329; 0.519]	0.660

SE: standard error; CI: confidence interval; NP: non parametric; Individual ad hoc methods (Indiv. nAUC): 1. Data: raw data, 2. LOCF: last observation carried forward, 3.Mean Imp.: mean imputation.

treatment is calculated with fixed parameter estimates, with the traditional ones where the nAUC is calculated using the trapezoidal method for every individual and compared at group level with a two-sample t-test. Similarly to the study of simulated data, estimates of individual nAUCs are computed using either log-transformed raw data without any transformation, LOCF or mean imputation ad hoc approaches. In addition, we applied the non-parametric approach NP nAUC briefly defined in section 4, Simulation study. The results are gathered in Table 4. In vac-IL2, the proposed approach concluded a significant difference between the two groups of treatment with a p-value of 0.031. Similar result is obtained with raw data with p-value slightly lower than 0.05. However, both LOCF, mean imputation ad hoc methods and non-parametric method are unable to reject the null hypothesis. All the tests lead to the same conclusion of no difference between groups in the LIGHT study. Considering the mean trajectories of the control and vaccine groups displayed in Figure 4(a) and (b), all the results obtained with our new approach are consistent with expected conclusions.

6 Discussion

In this paper, we proposed a splines-MEM based approach to estimate and compare the normalized area under the longitudinal outcome curve when observations are subject to left-censoring, induced by an LOD, and MAR monotonic missing data, due to drop-out. We demonstrated in a simulation study that incomplete data leads to biased estimates of nAUC resulting in invalid inferences regarding the difference in nAUC between groups with individual methods even when using simple ad hoc missing data correction, such as LOCF and mean imputation. Compared to the latter, we illustrated the superiority of our approach in term of type-I error and power. In addition, although the non-parametric approach developed by Vardi et *al.*[25] provided as robust statistical properties as our proposed method while the percentages of left-censored data remained lower than 50%, corresponding to a threshold of ART resumption higher than 100,000 copies/ml, the lack of information induced by higher percentages of drop out resulted in weaker results under certain conditions of simulation and more biased estimations of the difference of nAUC. We also highlighted that when the amount of data with drop-out is as high as 80% such as in a situation when ART are resumed if HIV RNA viral load exceeds 10,000 copies/mL in ATI trial, only the parametric approach appeared efficient to compare nAUC between groups. An application of two ATI trials for HIV illustrates the superiority of our method on real data.

Limitations of the proposed method include some assumptions induced by the use of MEM such as the normality and the homoscedasticity. However, we demonstrated that on clinical data these assumptions are realistic. As briefly noticed in section 3 (*Method*), two other versions of the proposed method are presented in Online Appendix replacing the estimation of ΔnAUC through the most commonly used trapezoid method by its estimation with either Lagrange or Spline interpolation methods. No significant differences of robustness have been observed in the application of those three methods on our well defined and tightened simulated trial designs. However, Lagrange and Splines methods could present more robust results in case of sparse designs. Also, in our simulations, we assumed a balanced longitudinal design with equal number of measurements and constant time points for every subject. Although clinical trials are commonly designed with the same monitoring for all participants, in reality the observed follow-up may deviate from the expected one. Moreover, some clinical trials could be designed to compare different monitoring designs among group in addition to treatment efficacy. As defined, the proposed method, being based on a discrete method of AUC calculation, should be biased by unbalanced times of measurements among groups with varying number of time points as well as different and irregular time steps between groups. As mentioned by Chandrasekhar et al.,[18] the consideration of time as continuous variable

in the AUC calculation could be a solution to handle this problem. To this end, we could either refine the time grid to mimic continuous time in the AUC calculation step, or use more complex AUC approximation methods such as Gaussian quadrature methods. The choice of Gaussian quadrature methods requires thus the use of a resampling procedure, such as bootstrapping to estimate the standard error. In clinical trials, the sample size calculation, resulting in the determination of the number of participants in each arm needed to detect a clinically relevant treatment effect, is one of the major steps in designing the study. The proposed statistics being defined as classical Z-statistics, typical formulas of sample size calculation can be derived from it. As defined by Hazra et al.,[47] the general formula for two-sided test can be given by $n = (Z_{1-\alpha/2} + Z_{1-\beta})^2 \times \sigma^2/\delta^2$ where α represents the accepted type-I error, β the type-II error, σ the standard deviation of the outcome being studied and δ the size effect defined as the targeted $\Delta n\text{AUC}/2$ in our case. Adjusted formulas can also be derived from this latter to account for unequal sized groups or unequal variance of outcomes using pooled variances. Simulations can be found in Online Appendix (see Figure F.1 in Online Appendix F) and showed good concordance between theoretical and practical power when there is no missing data. When missing data arise due to left censoring (LOD) or informative drop out, one need to take it into account in the sample size calculation.

The simulation study has been led under model correct specification assumption, i.e. the model used to analyse the data corresponds to the true data generation process. We further relaxed this assumption by using adaptive splines model for which some parameters, such as the location and number of knots for splines are supposed unknown.

Various extensions of this work could be guided to address the problem when there are a high proportion of drop-outs. The incorporation of prior information could be done through several ways. The study of more constrained splines through the addition of penalty on spline coefficients (P-splines)[48] or monotony and boundary conditions[49] (natural splines) is an option. In the same perspective, future research aims to extend this method to the use of mechanistic models.[50] In addition to introducing biological interpretation of the parameters, these models could incorporate more easily additional information such as asymptotic behaviors with steady states.

Acknowledgements

The authors thank the Vaccine Research Institute and the ANRS as sponsors and the LIGHT and VAC-IL2 study groups for sharing the data used in this article. Numerical computations were in part carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d'Aquitaine (see https://www.plafrim.fr/). The authors thank Dr Torsten Hothorn and two anonymous reviewers for constructive comments on the manuscript. The research has been initiated in the contest of the EHVA T01 trial which is supported by the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 681032) and the Swiss Government (grant number 15 0337).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Additional information

Web Appendix is available with this paper at the Statistical Methods in Medical Research website. R Code implementing the method is available on github at https://github.com/marie-alexandre/AUCcomparison.git. A reference manual has been included in the package (https://github.com/marie-alexandre/AUCcomparison/blob/master/Reference_manual.pdf) describing how to implement the proposed method.

ORCID iD

Marie Alexandre https://orcid.org/0000-0002-3557-7075

References

1. Scheff JD, Almon RR, DuBois DC, et al. Assessment of pharmacologic area under the curve when baselines are variable. *Pharmaceut Res* 2011; **28**: 1081–1089.

2. Heldens J, Weststrate M and Van den Hoven R. Area under the curve calculations as a tool to compare the efficacy of equine influenza vaccines – a retrospective analysis of three independent field trials. *J Immunol Methods* 2002; **264**: 11–17.
3. Lydick E, Epstein R, Himmelberger D, et al. Area under the curve: a metric for patient subjective responses in episodic diseases. *Qual Life Res* 1995; **4**: 41–45.
4. Neoptolemos JP, Stocken DD, Friess H, et al. A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer. *N Engl J Med* 2004; **350**: 1200–1210.
5. Schleyer E, Kühn S, Rührs H, et al. Oral idarubicin pharmacokinetics – correlation of trough level with idarubicin area under curve. *Leukemia* 1996; **10**: 707–712.
6. Duh MS, Lefebvre P, Fastenau J, et al. Assessing the clinical benefits of erythropoietic agents using area under the hemoglobin change curve. *The Oncologist* 2005; **10**: 438–448.
7. Hothorn LA. Statistical analysis of in vivo anticancer experiments: tumor growth inhibition. *Drug Inform J* 2006; **40**: 229–238.
8. Wu J and Houghton PJ. Interval approach to assessing antitumor activity for tumor xenograft studies. *Pharmaceut Stat* 2010; **9**: 46–54.
9. Qian W, Parmar M, Sambrook R, et al. Analysis of messy longitudinal data from a randomized clinical trial. *Stat Med* 2000; **19**: 2657–2674.
10. Cole SR, Napravnik S, Mugavero MJ, et al. Copy-years viremia as a measure of cumulative human immunodeficiency virus viral burden. *Am J Epidemiol* 2010; **171**: 198–205.
11. Ramos EL, Mitcham JL, Koller TD, et al. Efficacy and safety of treatment with an anti-m2e monoclonal antibody in experimental human influenza. *J Infect Dis* 2015; **211**: 1038–1044.
12. Calfee DP, Peng AW, Cass LM, et al. Safety and efficacy of intravenous zanamivir in preventing experimental human influenza a virus infection. *Antimicrob Agents Chemother* 1999; **43**: 1616–1620.
13. Allison DB, Paultre F, Maggio C, et al. The use of areas under curves in diabetes research. *Diabetes Care* 1995; **18**: 245–250.
14. Venter C, Slabber M and Vorster H. Labelling of foods for glycaemic index-advantages and problems. *South African J Clin Nutr* 2003; **16**: 118–126.
15. Potteiger J, Jacobsen D and Donnelly J. A comparison of methods for analyzing glucose and insulin areas under the curve following nine months of exercise in overweight adults. *Int J Obesity* 2002; **26**: 87–89.
16. Wilding GE, Chandrasekhar R and Hutson AD. A new linear model-based approach for inferences about the mean area under the curve. *Stat Med* 2012; **31**: 3563–3578.
17. Bell ML, King MT and Fairclough DL. Bias in area under the curve for longitudinal clinical trials with missing patient reported outcome data: summary measures versus summary statistics. *SAGE Open* 2014; **4**: 2158244014534858.
18. Chandrasekhar R, Shi Y, Hutson AD, et al. Likelihood-based inferences about the mean area under a longitudinal curve in the presence of observations subject to limits of detection. *Pharmaceut Stat* 2015; **14**: 252–261.
19. Little R and An H. Robust likelihood-based analysis of multivariate data with missing values. *Stat Sin* 2004; **14**: 949–968.
20. Jacqmin-Gadda H, Thiébaut R, Chêne G, et al. Analysis of left-censored longitudinal data with application to viral load in HIV infection. *Biostatistics* 2000; **1**: 355–368.
21. Thiébaut R and Jacqmin-Gadda H. Mixed models for longitudinal left-censored repeated measures. *Comput Meth Progr Biomed* 2004; **74**: 255–260.
22. Vaida F and Liu L. Fast implementation for normal mixed effects models with censored response. *J Comput Graph Stat* 2009; **18**: 797–817.
23. Schisterman E and Rotnitzky A. Estimation of the mean of a k-sample u-statistic with missing outcomes and auxiliaries. *Biometrika* 2001; **88**: 713–725.
24. Spritzler J, DeGruttola VG and Pei L. Two-sample tests of area-under-the-curve in the presence of missing data. *Int J Biostat* 2008; **4**.
25. Vardi Y, Ying Z and Zhang CH. Two-sample tests for growth curves under dependent right censoring. *Biometrika* 2001; **88**: 949–960.
26. Garner SA, Rennie S, Ananworanich J, et al. Interrupting antiretroviral treatment in HIV cure research: scientific and ethical considerations. *J Virus Eradicat* 2017; **3**: 82.
27. Li JZ, Etemad B, Ahmed H, et al. The size of the expressed HIV reservoir predicts timing of viral rebound after treatment interruption. *AIDS* 2016; **30**: 343.
28. Henderson GE, Peay HL, Kroon E, et al. Ethics of treatment interruption trials in HIV cure research: addressing the conundrum of risk/benefit assessment. *J Med Ethics* 2018; **44**: 270–276.
29. Sneller MC, Justement JS, Gittens KR, et al. A randomized controlled safety/efficacy trial of therapeutic vaccination in HIV-infected individuals who initiated antiretroviral therapy early in infection. *Sci Transl Med* 2017; **9**: eaan8848.
30. Sued O, Ambrosioni J, Nicolás D, et al. Structured treatment interruptions and low doses of il-2 in patients with primary HIV infection. inflammatory, virological and immunological outcomes. *PLoS One* 2015; **10**: e0131651.

31. Lévy Y, Thiébaut R, Montes M, et al. Dendritic cell-based therapeutic vaccine elicits polyfunctional hiv-specific t-cell immunity associated with control of viral load. *Eur J Immunol* 2014; **44**: 2802–2810.
32. Pollard RB, Rockstroh JK, Pantaleo G, et al. Safety and efficacy of the peptide-based therapeutic vaccine for hiv-1, vacc-4×: a phase 2 randomised, double-blind, placebo-controlled trial. *Lancet Infect Dis* 2014; **14**: 291–300.
33. Bar KJ, Sneller MC, Harrison LJ et al. Effect of HIV antibody vrc01 on viral rebound after treatment interruption. *N Engl J Med* 2016; **375**: 2037–2050.
34. Fagard C, Le Braz M, Günthard H, et al. A controlled trial of granulocyte macrophage-colony stimulating factor during interruption of HAART. *AIDS* 2003; **17**: 1487–1492.
35. Palich R, Ghosn J, Chaillon A, et al. Viral rebound in semen after antiretroviral treatment interruption in an HIV therapeutic vaccine double-blind trial. *AIDS* 2019; **33**: 279–284.
36. Lévy Y, Gahéry-Ségard H, Durier C, et al. Immunological and virological efficacy of a therapeutic immunization combined with interleukin-2 in chronically hiv-1 infected patients. *AIDS* 2005; **19**: 279–286.
37. Brundage TM, Vainorius E, Chittick G, et al. Brincidofovir decreases adenovirus viral burden, which is associated with improved mortality in pediatric allogeneic hematopoietic cell transplant recipients. *Biol Blood Marrow Transplant* 2018; **24**: S372.
38. Hill JA, Mayer BT, Xie H, et al. Kinetics of double-stranded DNA viremia after allogeneic hematopoietic cell transplantation. *Clin Infect Dis* 2018; **66**: 368–375.
39. Zecca M, Wynn R, Dalle JH, et al. Association between adenovirus viral load and mortality in pediatric allo-HCT recipients: the multinational advance study. *Bone Marrow Transplant* 2019; **54**: 1632–1642.
40. Kosulin K, Pichler H, Lawitschka A, et al. Diagnostic parameters of adenoviremia in pediatric stem cell transplant recipients. *Front Microbiol* 2019; **10**: 414.
41. Satterthwaite FE. An approximate distribution of estimates of variance components. *Biometr Bull* 1946; **2**: 110–114.
42. Hrong-Tai Fai A and Cornelius PL. Approximate f-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *J Stat Comput Simulat* 1996; **54**: 363–378.
43. Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997; **53**: 983–997.
44. Bailer AJ. Testing for the equality of area under the curves when using destructive measurement techniques. *J Pharmacokinet Biopharmaceut* 1988; **16**: 303–309.
45. Spiriti S, Smith P and Lecuyer P. *freeknotsplines*: *algorithms for implementing free-knot splines*, 2018, R package version 1.0.1, https://cran.r-project.org/web/packages/freeknotsplines/index.html.
46. Vaida F and Liu L. *lmec*: *linear mixed-effects models with censored responses*, 2012. . R package version 1.0, https://CRAN.R-project.org/package = lmec (accessed 8 June 2021).
47. Hazra A and Gogtay N. Biostatistics series module 5: determining sample size. *Ind J Dermatol* 2016; **61**: 496.
48. Eilers PH and Marx BD. Flexible smoothing with b-splines and penalties. *Stat Sci* 1996; **11**: 89–102.
49. Laurini MP and Moura M. Constrained smoothing b-splines for the term structure of interest rates. *Insurance* 2010; **46**: 339–350.
50. Perelson AS and Ribeiro RM. Introduction to modeling viral infections and immunity. *Immunol Rev* 2018; **285**: 5.
51. Berrut JP and Trefethen LN. Barycentric Lagrange interpolation. *SIAM Rev* 2004; **46**: 501–517.

6.5 Clairon et al. 2023 (Plos Comp. Biol.) Modeling Humoral Response in COVID-19

Modeling the kinetics of the neutralizing antibody response against SARS-CoV-2 variants after several administrations of Bnt162b2. Clairon Q., Prague M. Planas D., Bruel T., Hocqueloux L., Prazuck T., Schwartz O., Thiébaut R. and Guedj J. *PLOS Computational Biology.* 19(8) - e1011282 - August 2023.

This article integrates in the second axis of my research "Within-host Modeling in Infectious Diseases". It has been written by my postdoctoral fellow and myself in collaboration with Jérémie Guedj from Inserm Paris IAME.

I selected this work because it exemplifies the mechanistic modeling of immune dynamics. Furthermore, it demonstrates how mechanistic models can be utilized for their predictive capabilities, specifically in this case to assess the waning immunity after COVID-19 vaccination. One of the main challenges in this work was the identifiability of the model when only antibody data are observed in a relatively sparse framework.

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Modeling the kinetics of the neutralizing antibody response against SARS-CoV-2 variants after several administrations of Bnt162b2

Quentin Clairon[1,2,3☯¤], **Mélanie Prague**[1,2,3☯¤], **Delphine Planas**[3,4], **Timothée Bruel**[3,4‡], **Laurent Hocqueloux**[5‡], **Thierry Prazuck**[5‡], **Olivier Schwartz**[3,4‡], **Rodolphe Thiébaut**[1,2,3‡*], **Jérémie Guedj**[6]

1 Université de Bordeaux, Inria Bordeaux Sud-Ouest, Bordeaux, France, **2** Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR1219, Bordeaux, France, **3** Vaccine Research Institute, Créteil, France, **4** Virus and Immunity Unit, Institut Pasteur, Université de Paris Cité, CNRS UMR3569, Paris, France, **5** Service des Maladies Infectieuses et Tropicales, Centre Hospitalier Régional, Orléans, France, **6** Université Paris Cité, IAME, Inserm, Paris, France

☯ These authors contributed equally to this work.
¤ Current address: Bordeaux Population Health Research Center, Université de Bordeaux, 146 rue Léo Saignat,33076 Bordeaux cedex, France
‡ TB, LH, TP and OS also contributed equally to this work.
* rodolphe.thiebaut@u-bordeaux.fr

OPEN ACCESS

Citation: Clairon Q, Prague M, Planas D, Bruel T, Hocqueloux L, Prazuck T, et al. (2023) Modeling the kinetics of the neutralizing antibody response against SARS-CoV-2 variants after several administrations of Bnt162b2. PLoS Comput Biol 19(8): e1011282. https://doi.org/10.1371/journal.pcbi.1011282

Editor: James R. Faeder, University of Pittsburgh, UNITED STATES

Received: December 19, 2022

Accepted: June 20, 2023

Published: August 7, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1011282

Data Availability Statement: All relevant data are within the manuscript and its Supporting information files.

Abstract

Because SARS-CoV-2 constantly mutates to escape from the immune response, there is a reduction of neutralizing capacity of antibodies initially targeting the historical strain against emerging Variants of Concern (VoC)s. That is why the measure of the protection conferred by vaccination cannot solely rely on the antibody levels, but also requires to measure their neutralization capacity. Here we used a mathematical model to follow the humoral response in 26 individuals that received up to three vaccination doses of Bnt162b2 vaccine, and for whom both anti-S IgG and neutralization capacity was measured longitudinally against all main VoCs. Our model could identify two independent mechanisms that led to a marked increase in measured humoral response over the successive vaccination doses. In addition to the already known increase in IgG levels after each dose, we identified that the neutralization capacity was significantly increased after the third vaccine administration against all VoCs, despite large inter-individual variability. Consequently, the model projects that the mean duration of detectable neutralizing capacity against non-Omicron VoC is between 348 days (Beta variant, 95% Prediction Intervals PI [307; 389]) and 587 days (Alpha variant, 95% PI [537; 636]). Despite the low neutralization levels after three doses, the mean duration of detectable neutralizing capacity against Omicron variants varies between 173 days (BA.5 variant, 95% PI [142; 200]) and 256 days (BA.1 variant, 95% PI [227; 286]). Our model shows the benefit of incorporating the neutralization capacity in the follow-up of patients to better inform on their level of protection against the different SARS-CoV-2 variants.

Funding: Work in OS lab is funded by Institut Pasteur, Urgence COVID-19 Fundraising Campaign of Institut Pasteur, Fondation pour la Recherche Médicale (FRM), ANRS, the Vaccine Research Institute (ANR-10-LABX-77), Labex IBEID (ANR-10-LABX-62-IBEID), ANR / FRM Flash Covid PROTEO-SARS-CoV-2, ANR Coronamito, HERA european funding, Sanofi and IDISCOVR. This work has received funding from the French Agency for Research on AIDS and Emerging Infectious Diseases via the EMERGEN project (ANRS0151). This work was supported by INSERM and the Investissements d'Avenir program, Vaccine Research Institute (VRI), managed by the ANR under reference ANR-10-LABX-77-01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors declare no competing interests.

Trial registration: This clinical trial is registered with ClinicalTrials.gov, Trial IDs NCT04750720 and NCT05315583.

Author summary

Developed vaccines against SARS-CoV-2 have been a turning point against the ongoing Covid-19 pandemic. When the Wuhan virus was dominant, they help to dramatically reduce the number of severe cases as well as infection and transmission rates. For mRNA vaccines, it was in great part explained by the high level of induced antibodies a few weeks/months after injection and linked to high neutralizing capacity, the ability to prevent viruses to enter and infect target cells. However, decreasing antibody concentration over time and apparition of variants escaping their neutralizing action dramatically reduced the initial vaccine efficacy. As a countermeasure, additional injections were used to re-establish significant antibody population and ensure a long-term neutralizing activity against emerging variants. To infer if this multi-dose strategy fulfills such task, we construct a model of the evolution of the induced antibodies and their neutralizing capacity against different variants. This model helps us to quantify the gain brought by each new injection on both antibody population and their neutralizing ability against all tested variants as well as the dramatic differences between them. We also predict the long-term evolution of neutralizing activity, years after last injection, and thus discuss the longevity of the induced protection by vaccine.

Introduction

The discovery and the rapid availability of several vaccines against SARS-CoV-2 has been a turning point in the combat against Covid-19 [1]. Although their efficacy may vary to some extent, it is undisputable that large scale vaccination campaigns have dramatically reduced both the risk of severe diseases [2–4] and, to a lesser extent, the rates of transmission and disease acquisition [5–7], resulting in millions of saved lives [1, 8, 9].

However vaccine efficacy has been jeopardized by the apparition of various Variants of Concern (VoCs) that partially escape immune protection. A clear decrease in the neutralization capacity has been observed [10, 11] which has translated to a substantial reduction of efficacy against transmission and disease acquisition with Delta and Omicron variants, and, to a lesser extent, to a decrease of efficacy against severe Covid-19 disease [12, 13]. The concern caused by a potential loss of protection against VoCs has been further enhanced by the natural waning immunity and the progressive reduction in antibody levels over time [14–16]. This has supported boosting strategies with one or two additional vaccine doses to maintain a high level of protection. However the optimal time to administer boosters, and how these times may vary for different VoCs, remains unclear.

To characterize in detail the duration of protection against SARS-CoV-2, it is therefore essential to measure not only total anti-S IgG antibodies over time, as typically done in large observational studies, but also how this translates in terms of neutralization capacity. The latter requires intensive *in vitro* measurements, but it provides a much more accurate description of the level of protection present in the sera of Covid-19 vaccine recipients [17, 18]. Then, a detailed characterization of the immunological or virological factors modulating the

duration of protection can be obtained by using mathematical models of immune marker dynamics [19].

Here we propose to use for the first time a mathematical model to analyze the joint kinetics of anti-S IgG antibodies and neutralization capacity after repeated vaccine injections against the main VoCs. For that purpose we relied on data from a cohort of Bnt162b2 vaccine recipients, in which both antibody kinetics and neutralizing activity were measured longitudinally [11, 20, 21]. We built on previous models of antibody kinetics [22, 23] to develop a novel approach to quantify the kinetics of neutralizing activity, and we use this model to characterize the effects of repeated vaccine administrations on it. We finally use the model to discuss the duration of protection conferred by the measured humoral activity induced by Bnt162b2 against VoCs.

Materials and methods

Ethics statement

This study was approved by the Ethics Committee ILE DE FRANCE IV. The cohort was approved by the national external committee (CPP Ile-de-France- IV IRB No. 00003835). Study participants did not receive any compensation. At enrolment a written informed consent was collected for all participants.

Data

Population study. Data originate from a cohort of N = 29 subjects who received up to three injections of Bnt162b2 (ClinicalTrials.gov:NCT04750720 and ClinicalTrials.gov: NCT05315583). In brief SARS-CoV-2 naive patients were recruited in Orléans, France between August 27, 2020 and May 24, 2022. Individuals were followed for up to 483 days after their first vaccine injections (see more details on the data in [11, 20, 21]). Two patients without longitudinal follow-up and 1 immunocompromised individual were not included in our analysis. In total, $N = 26$ individuals were analyzed (see Table 1). Briefly, all subjects received at least 2 doses, administered on average 27 days after the first injection. $N = 22$ subjects received a third injection, administered on average 269 days after the first injection. During the follow-up $N = 12$ had a positive PCR, and only data prior to infection were analyzed, leaving an average follow-up of 11 visits and a median follow-up time of 362 days.

Longitudinal markers of immune response. Two types of measurements were available at each visit: 1) anti-spike binding IgGs, measured in BAU/mL) neutralization titers of sera

Table 1. Characteristics of the analyzed population.

Characteristics	Median [Min; Max] or n (%)	Median Time of vaccination [Min; Max]	
		since first dose	since second dose
Men	14 (54%)		
Age	59 [33; 95]		
Follow-up duration after first-dose (days)	368 [168; 483]		
Number of follow-up visits	14 [2; 18]		
Number of vaccination doses			
1st	26 (100%)	-	-
2nd	26 (100%)	22 [17; 60]	-
3nd	23 (88%)	243 [175; 385]	221 [154; 361]

https://doi.org/10.1371/journal.pcbi.1011282.t001

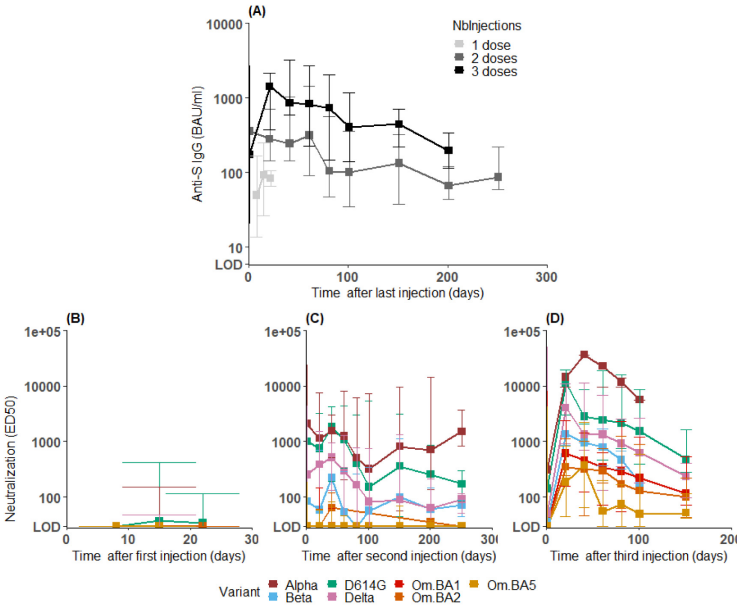


Fig 1. A: longitudinal evolution of the binding antibody concentration of anti-S IgG. B-C-D: longitudinal evolution of the neutralizing activity against VoCs after the first (B, see S1 Fig for a zoomed version), second (C) and third (D) vaccination dose. Squares represent median values, and plain horizontal lines represent the minimal and maximal encountered values among subjects. The lower limit of detection (LOD) is equal to 6 BAU/mL for IgG and 30 for ED_{50}. Given the limited number of samples available, data were grouped, using a one week sliding window after the first dose, 20 days in the first 100 days following the second or third infection, and 50 days for the other data points.

https://doi.org/10.1371/journal.pcbi.1011282.g001

provided in ED_{50}, which is the effective dilution required to neutralize 50% of an arbitrary viral load of reference (eg, the higher the ED_{50} the larger the protection level). Neutralization capacity was assessed against historical strain (D614G), Alpha, Beta, Delta, and Omicron variants (strains BA.1, BA.2, and BA.5).

In brief IgGs markedly increased after each dose, but rapidly declined over time, with a rate that did not substantially differ after the second or the third dose (Fig 1A). In contrast the kinetics of neutralizing activity was much more heterogeneous, and was characterized by large differences against the different VoCs. Further the neutralizing activity was markedly increased after the third dose against all Omicron variants, albeit remaining at much lower levels than against the other VoCs (Fig 1B, 1C and 1D).

Model for neutralizing antibody response

Mechanistic model for antibody kinetics. We rely on a simplified and rescaled version of a previously published model in the context of vaccine against Ebola infection [23]. In brief, after each dose, cells transfected with Bnt162b2 generate antigen, noted V, which triggers the constitution of a memory compartment, noted M, at a rate ρ. This memory compartment is a general one accounting for all cell populations able to differentiate into secreting cells upon antigen presence. These can be activated or memory B-cells either circulating or present in

germinal centers. So, M can differentiate into secreting plasma cells, noted $\tilde{S}$, at a rate μV. These cells then produce antibodies, noted Ab, at a rate θ. V, S, and Ab are degraded at rates δ_V, δ_S, and δ_{Ab}, respectively, leading to the following ODE system:

$$\begin{aligned} \dot{V} &= -\delta_V V \\ \dot{M} &= \rho V - \mu V M \\ \dot{\tilde{S}} &= \mu V M - \delta_S \tilde{S} \\ \dot{Ab} &= \theta \tilde{S} - \delta_{Ab} Ab. \end{aligned} \tag{1}$$

Assuming that individuals are naive of infection, and noting $\mathbf{t_1}$ the time of first injection, the initial conditions are given by: $M(\mathbf{t_1}) = \tilde{S}(\mathbf{t_1}) = Ab(\mathbf{t_1}) = 0$.

To model the effect of repeated doses, we consider that V is a function presenting discontinuity at time of first, second and third injection ($\mathbf{t_1}$, $\mathbf{t_2}$, $\mathbf{t_3}$). By denoting $k = 1, \ldots, 3$ the dose number, on each interval $[\mathbf{t_k}, \mathbf{t_{k+1}}]$, solving previous ODE for V gives us $V_k(t) = V_0 e^{-\delta_V (t - \mathbf{t_k})}$ where V_0 is the initial antigen concentration, assumed equal from all doses.

Because this model is not identifiable when only Ab are measured, we derived a structurally identifiable approximated model described in Eq 2; see Appendix A.1 in S1 Appendix for a description of this simplification. Briefly, it consists of rescaling the model for $S = (\mu V_0 \overline{M_1})^{-1} \tilde{S}$ and assuming that M can be replaced by its steady-state value $\overline{M_k}$ if equilibrium is reached quickly after each injection:

$$\begin{aligned} \dot{S} &= f_{\overline{M_k}}\, e^{-\delta_V (t - \mathbf{t_k})} - \delta_S S \\ \dot{Ab} &= \vartheta S - \delta_{Ab} Ab \end{aligned} \tag{2}$$

where $f_{\overline{M_k}} = \frac{\overline{M_k}}{\overline{M_1}}$ is the fold-change for steady-state memory compartment after k^{th} injection compared to the first one (by definition $f_{\overline{M_1}} = 1$). Of note, we also tested the full model which does not assume a steady state value for M. This leads to identifiability issues mainly due to μ_S estimation for which only a lower bound ($\mu_S > 20$) can be found. For such values for μ_S, the compartment M nearly instantaneously reaches its steady-state. Accordingly, both full and simplified models provide virtually similar predictions for Ab (see Appendix A.2 in S1 Appendix). Finally, we also tested a more complex model accounting for a delay between vaccine injection and antibody production (also in Appendix A.2 in S1 Appendix). However the model did not improve data description, which was probably due to the limited amount of information available on antibody kinetics in the couple of days following vaccine injection. Moreover, the model proposed by Balelli et al. [23] initially contains two populations of secreting cells S and L, differing by their life expectancy. In our case, preliminary statistical analysis conclude that there was no statistical differences between model adjustments when accounting for S and L or S only (results not shown). This allows us to reduce the number of unknown parameters. This is crucial for parameters related to cell kinetics known to be very different for newly developed mRNA vaccines comparing to viral vector ones and for which no values have been previously inferred. Thus, the retained model (2) is complex enough to account for the effect of multiple injections on antibody concentration evolution while avoiding identifiability issues. We define $\eta^{ODE} = (f_{\overline{M_2}}, f_{\overline{M_3}}, \delta_V, \delta_S, \vartheta, \delta_{Ab})$ the vector of model parameters defining the dynamics of the system.

Functional model for neutralizing activity. After modeling antibody concentration evolution in the previous section, we aim to model their neutralizing activity. This means in our

case proposing a model describing the evolution of ED_{50}^{v} with respect to Ab. We consider the following linear model:

$$ED_{50}^{v}(t) = F(v,t)Ab(t).$$

The function $F(v, t)$ represents the relationship between the concentration of binding antibodies in BAU/mL and its neutralization capacity against the VoC v. It is variant-specific and time-varying, let us first derive its expression for the strain D614G before moving to any arbitrary VoCs. After $\mathbf{t_1}$, we assume a proportional relationship between Ab and neutralizing activity against D614G i.e $F(D614G, t) = \gamma$ (equivalently $ED_{50}^{D614G}(t) = \gamma Ab(t)$). After additional injections, we assume there is a neutralization gain quantified by the fold-change f_2 after $\mathbf{t_2}$ and f_3 after $\mathbf{t_3}$ i.e. $F(D614G, t) = \gamma f_2$ when $t \in [\mathbf{t_2};\mathbf{t_3}]$ and $F(D614G, t) = \gamma f_3$ for $t \geq \mathbf{t_3}$. Now, we account for VoCs specific neutralizing activity by modifying baseline value γ by the fold-changes f_v such that $F(v, t) = F(D614G, t)f_v = \gamma f_v$ when $t \in [\mathbf{t_1};\mathbf{t_2}]$ and $F(v, t) = F(D614G, t)f_v = \gamma f_v f_2$ for $t \in [\mathbf{t_2};\mathbf{t_3}]$. We assume that the relative gain brought by third injection can be also VoC-specific. That is why we introduce the fold-changes g_v to quantify this gain i.e. $F(v, t) = F(D614G, t)f_v g_v = \gamma f_v f_3 g_v$ for $t \geq \mathbf{t_3}$. This piece-wise constant function can be then expressed in a general form:

$$F(v,t) = \gamma f_v(\mathbb{1}_{t<\mathbf{t_2}} + f_2\mathbb{1}_{t\in[\mathbf{t_2};\mathbf{t_3}]} + f_3 g_v \mathbb{1}_{t\geq\mathbf{t_3}}).$$

The choice of this model is the result of exploration based on the minimization of an adjustment criteria. In particular, the current model only quantifies the effect of the repetition of injections on affinity enhancement. Other factors can play a role as the elapsed time since antigen presentation, for example to account for the progressive Memory B-cells repertoire expansion [24, 25]. An alternative neutralization model only considering the time factor has been developed. This supplementary analysis is described in Appendix B in S1 Appendix but lead to a less accurate model (in terms of AIC). A general model accounting for both factors, the number of injections and the elapsed time, has been also tested leading to non-significant improvements over the retained model and at the expense of identifiability problems (results not shown). We also investigate the possibility of a variant-specific fold-change after second injection. This was discarded due to practical identifiability issues. More generally, our model assumes a simple linear relationship between antibody concentration and neutralization, with no saturation effect. The fact that a more physiological model assuming a nonlinear relationship did not improve data description (see Appendix B in S1 Appendix) suggests that the level of Antibody observed in this study remains within the linear range of neutralization. Finally, we acknowledge the existence of other ways than our descriptive approach to link Ab and ED_{50}^{v}, as in Padmanabhan et al. [26] in which a mechanistic relationship between these quantities is constructed. Still, their model definition involves measurements, such as infection events or neutralizing antibodies, which are not at our disposal, especially for emerging VoCs, making their model intractable for our prediction purpose.

We define $\eta = (\eta^{ODE}, \gamma, f_v, f_2, f_3, g_v)$ the vector of model parameters that have to be estimated from the observed data. Description of the model parameters can be found in Table 2.

Observation model. The structural model used to describe the log-transformed concentration of binding antibodies in BAU/mL for the i^{th} individual ($i = 1, \ldots, N$) at the j^{th} time point ($j = 1, \ldots, n_i$) is:

$$Y_{ij}^{BAU} = \log_{10}(Ab(\eta_i, t_{ij})) + e_{ij}^{BAU},$$

where e_{ij}^{BAU} is the residual additive error which follows a normal distribution of mean zero and constant standard deviation σ_{BAU}. The vector η_i is the specific value for individual i of vector η.

Table 2. Model parameters and estimation. Fcn:fold-change in neutralization.

Parameter	Description	Unit	Fixed Effect [IC95%]	Random Effect [IC95%]
$f_{\overline{M_2}}$	Fold change for M equilibrium after second injection	dimensionless	7.1 [4.2; 12.0]	0.9 [0.8; 1.0]
$f_{\overline{M_3}}$	Fold change for M equilibrium after third injection	dimensionless	18.5 [15.0; 26.0]	
ϑ	Initial acceleration for Ab production	[A].days^{-2}	24.5 [15.8; 38.0]	0.5 [0.3; 0.7]
δ_{Ab}	Antibody degradation rate	days^{-1}	0.08 [0.07; 0.09]	
γ	Proportion of neutralization provided by first vaccination	[V].[A]$^{-1}$	0.3 [0.2; 0.5]	0.7 [0.5; 0.9]
f_{Alpha}	Fcn for variant Alpha	unitless	1.3 [1.0; 1.8]	
f_{Bseta}	Fcn for variant Beta	unitless	0.2 [0.1; 0.3]	
f_{Delta}	Fcn for variant Delta	unitless	0.3 [0.2; 0.4]	
$f_{BA.1}$	Fcn for variant BA.1	unitless	0.005 [0.003; 0.009]	
$f_{BA.2}$	Fcn for variant BA.2	unitless	0.013 [0.005; 0.029]	
$f_{BA.5}$	Fcn for variant BA.5	unitless	0.016 [0.011; 0.022]	
f_2	Fcn for second vs. first injection	unitless	8.2 [4.0; 16.9]	
f_3	Fcn for third vs. first injection in original strains D614G	unitless	18.8 [10.0; 42.9]	
g_{Alpha}	Fcn for third vs. first injection in variant Alpha	unitless	5.8 [3.0; 11.8]	
g_{Beta}	Fcn for third vs. first injection in variant Beta	unitless	2.3 [1.3; 4.2]	
g_{Delta}	Fcn for third vs. first injection in variant Delta	unitless	1.1 [0.6; 1.5]	
$g_{BA.1}$	Fcn for third vs. first injection in variant BA.1	unitless	13.5 [7.5; 24.3]	
$g_{BA.2}$	Fcn for third vs. first injection in variant BA.2	unitless	5.4 [2.5; 11.9]	
$g_{BA.5}$	Fcn for third vs. first injection in variant BA.5	unitless	1.7 [1.2; 2.5]	
σ_{BAU}	Measurement error for $Y^{BAU} = \log_{10}(Ab) + e^{BAU}$		0.24 [0.23; 0.25]	
σ_{D614G}	Measurement error for $Y^{D614G} = \log_{10}(ED_{50}^{D614G}) + e^{D614G}$		0.47 [0.44; 0.50]	
σ_{Alpha}	Measurement error for $Y^{Alpha} = \log_{10}(ED_{50}^{Alpha}) + e^{Alpha}$		0.59 [0.53; 0.64]	
σ_{Beta}	Measurement error for $Y^{Beta} = \log_{10}(ED_{50}^{Beta}) + e^{Beta}$		0.47 [0.41; 0.53]	
σ_{Delta}	Measurement error for $Y^{Delta} = \log_{10}(ED_{50}^{Delta}) + e^{Delta}$		0.42 [0.40; 0.44]	
$\sigma_{BA.1}$	Measurement error for $Y^{BA.1} = \log_{10}(ED_{50}^{BA.1}) + e^{BA.1}$		0.44 [0.36; 0.52]	
$\sigma_{BA.2}$	Measurement error for $Y^{BA.2} = \log_{10}(ED_{50}^{BA.2}) + e^{BA.2}$		0.48 [0.40; 0.56]	
$\sigma_{BA.5}$	Measurement error for $Y^{BA.5} = \log_{10}(ED_{50}^{BA.5}) + e^{BA.5}$		0.34 [0.30; 0.40]	
δ_V	Induced vaccine antigen declining rate	days^{-1}	2.7 (fixed)	
δ_S	Death rate of S cells	days^{-1}	0.01 (fixed)	

https://doi.org/10.1371/journal.pcbi.1011282.t002

We also consider a log-transformation of ED_{50}^{v} raw measurements for the variant in the list {*D*614*G*, *Alpha*, *Beta*, *Delta*, *BA*.1, *BA*.2, *BA*.5}. For the i^{th} individual at the j^{th} time point, we have:

$$Y_{ij}^{v} = \log_{10}(ED_{50}^{v}(\eta_i, t_{ij})) + e_{ij}^{v},$$

where e_{ij}^{v} is the residual additive error for variant v which follows a normal distribution of mean zero and constant standard deviation σ_v.

Statistical model for parameters over time and injections. **Fixed parameters**. Here, not all parameters can be jointly estimated via likelihood when only concentration of binding antibodies and antibody neutralizing activity are measured. Further, the model predictions were found largely insensitive to the choice of the degradation rate of V and S. Using a profiled likelihood approach [27], we fixed their half-life to 0.25 and 51 days, respectively.

Inter-individual variability. In the vector η, some parameters have to be individual-specific to account for inter-individual variability. It is the case for $\psi_i = (\vartheta, f_{\overline{M_2}}, \gamma)$. We suppose it

follows a log-normal distribution such that:

$$\psi_i = \psi_0 exp(u_i),$$

where ψ_0 is the fixed effect and average mean value in the population. The vector u_i is individual random effects, which follow a normal distribution of mean zero and standard deviation Ω, and account for heterogeneity across individual. We assume that other parameters in vector η except error measurements are also estimated in log-transformation and are common to all individuals in the population. Altogether, the vector of parameters to estimate is given by $\theta = (\eta, \Omega, \sigma_{BAU}, \sigma_v)$.

Estimation procedure. Parameters were estimated (and named $\hat{\theta}$ in the following) with the SAEM algorithm implemented in MONOLIX software version 2022R1 [28] allowing to handle left censored data [29]. Likelihood was estimated using the importance sampling method and standard error were obtained by asymptotic approximation and inversion of the Fisher Information Matrix. Graphical and statistical analyses were performed using R version 3.4.3.

Simulation of long-term humoral response. Next, we used the model to predict the long-term evolution of Ab and ED_{50}^v over time. To account for uncertainty in our predictions, we used a Monte-Carlo sampling method, where $K = 1000$ replicates of parameters values $\theta^{(k)}$ were sampled in the posterior distribution of the parameter estimates to derive 95% prediction intervals (PI) of the predicted trajectories.

Finally we used these predictions to calculate the time to reach a given threshold value. To take into account between-subjects variability, we added a second layer to our Monte-Carlo sampling method and we sampled $N = 100$ replicates in the population parameter distribution. We used these predictions to derive the probability of having a concentration of binding antibodies greater than given thresholds, in particular higher than 264 BAU/mL, which corresponds to the standard threshold of protection defined by Feng et al. [30] and adopted by WHO. The level of neutralizing activity has been identified as a correlate of protection for vaccine efficacy against the historical strain [31, 32]. However, to date, no threshold for ED_{50}^v value ensuring protection has been isolated for D614G, let alone for the new VoCs. So, for a range of threshold values, we calculated the probability that the neutralizing activity against each VoC would be higher than these values over time, especially if this activity was still detectable at a given time. In this way, we can compare the longevity of neutralizing activity between VoCs even in the absence of a clear threshold of protection for each of them.

Results

Mechanistic model for humoral response

We first aimed to investigate whether there is a proportional relationship between the evolution of concentration of binding antibodies and its neutralization capacity. Fig 2 displays the observed relationship from data between antibody concentration and ED_{50}^v for each VoC after each injection. First, we notice that these ratios are different for the variants. Then, we compared the evolution of these ratios with respect to the previous vaccination. In most cases, the ratios improved significantly, indicating an intrinsic gain in neutralization that cannot be explained by the variation in antibody concentration alone, justifying the need to quantify this phenomenon precisely. This is supported by the linear regressions of ED_{50}^v on Ab after each injection presented in Table 3 (made with the R package Censreg [33] to account for censored data). These regressions indicate an increased correlation between ED_{50}^v and Ab with respect to the injection numbers for most of VoCs (already pointed out by Goel et al. [34] for D614G and Beta).

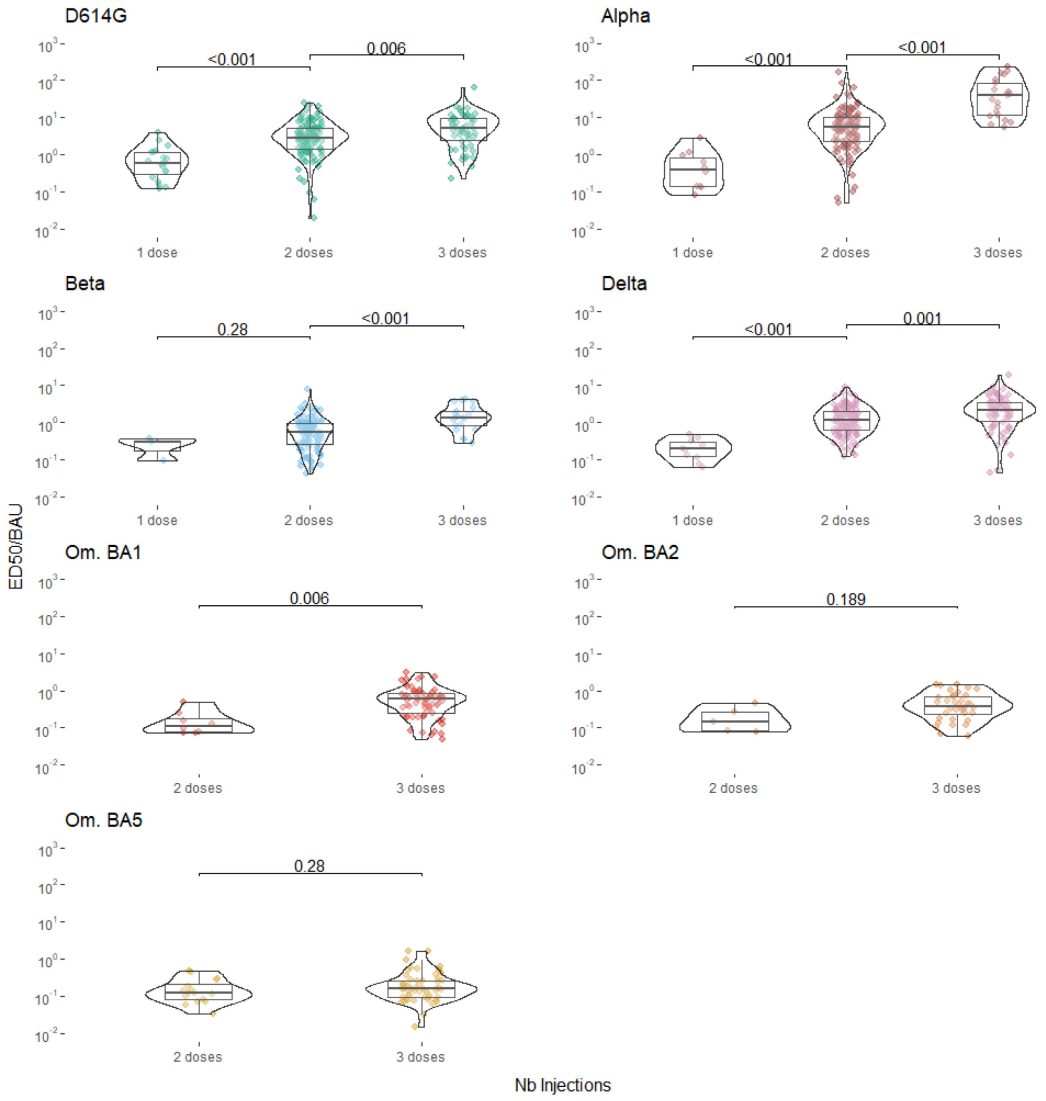


Fig 2. Evolution of the predicted ratio ED^{v}_{50}/BAU for each VoC after successive vaccine doses. Each circle represents a ratio ED^{v}_{50}/BAU computed when both measurements for ED^{v}_{50} and BAU where available for a given patient at a given observation time. Most of patients contribute several times due to the repeated measurements made over time after each dose. All predictions below the limit of detection for ED^{v}_{50} were removed to avoid overoptimistic ED^{v}_{50}/BAU ratio when replacing ED^{v}_{50} values by detection threshold. This explains why very few values are available for Beta and Delta and none for Omicron strains for one dose case. Comparison between vaccine dose was done using Wilcoxon test with Holm correction, p-values are given above the brackets.

https://doi.org/10.1371/journal.pcbi.1011282.g002

Table 3. Linear regression results with censored data.

	β estimation in model $ED_{50}^{v} = \alpha + \beta Ab$ on interval:		
	$[t_1;t_2]$	$[t_2;t_3]$	$[t_3;+\infty]$
D614G	0.90 [0.01; 1.73]	2.42 [1.48; 3.47]	15.31 [8.93; 21.68]
Alpha	0.50 [-0.26; 1.24]	4.39 [2.02; 6.76]	38.65 [3.24; 74.06]
Beta	0.10 [-0.15; 0.30]	0.30 [0.15; 0.45]	1.83 [0.93; 2.70]
Delta	2.05 [-10.5; 12.30]	1.58 [1.11; 2.02]	5.86 [4.16; 7.55]
BA.1	-	0.11 [0.01; 0.22]	1.08 [0.74; 1.41]
BA.2	-	0.16 [0.02; 0.30]	0.21 [0.08; 0.36]
BA.5	-	0.15 [0.07; 0.23]	0.51 [0.31; 0.71]

https://doi.org/10.1371/journal.pcbi.1011282.t003

Estimation of model parameters can be found in Table 2. This estimation indicates that multiple injections both increase antibody concentration and intrinsic affinity per constant antibody unit. Regarding antibody concentration, estimation of mechanistic parameters indicates a significant increase in the size of the memory compartment. It increased by $f_{\overline{M_2}} = 7.1$ (95% Confidence interval CI [4.2; 12.0]) after the second injection and by $f_{\overline{M_3}} = 18.5$ (95% CI [15.0; 26.0]) after the third injection compared to the first one. Of note, the estimated value for δ_{Ab} approximately corresponds to an half-life of 9 days, which is close to the typical range of 10 and 21 days [35].

Regarding neutralization per constant antibody concentration unit, we found that there are two main influencing factors: the repetition of the injections and the VoC. Regarding repeated injections effect for the original strain D614G, the second dose increases neutralization by a factor f_2 = 8.2 (95% CI [4.0; 16.9]) and the third one by f_3 = 18.8 (95% CI [10.0; 42.9]) compared to the first injection. Now regarding the neutralization capacities for emerging VoCs, they are significantly decreased compared to the original strain, with the exception of Alpha, where there is no significant change in neutralization compared to D614G. It ranges from a reduction of 70% (95% CI [60%; 80%]) for Delta to a dramatic reduction of 99.5% (95% CI [99.1%; 99.7%]) for BA.1. Still, we find that the sequential injection strategy confers a gain in long-term neutralizing capacities for all VoCs. The second injection increases neutralization against all VoCs by the same factor f_2 = 8.2 (same as D614G). The third injection increases neutralization in a VoC-specific manner, given by $f_3 g_v$. It ranges from an increase in fold change of 21 (95% CI [6.0; 64.4]) for Delta to 254 (95% CI [75; 1042]) for BA.1 times higher for the third injection than for the first injection. For comparison with D614G, the neutralization is $\frac{f_3}{f_2} = 2.3$ (95% CI [1.6; 3.2]) times higher for the third injection compared to second injection. Transitively, the fold change is $\frac{f_3}{f_2} g_{Delta} = 2.5$ (95% CI [0.8; 4.8]) for Delta to $\frac{f_3}{f_2} g_{BA.1} = 31.1$ (95% CI [12.0; 77.8]) for BA.1 times higher for the third injection than for the second injection.

Examples of fitted trajectories are given for four randomly selected patients in Fig 3. We observe a very good adequation with most of the observations lying in the 95% prediction intervals. To assess the capability of the model to fit our data, we also examined the visual predictive check (see Appendix C in S1 Appendix), which showed that the model well captures the kinetics observed and its variability across individuals.

This is exemplified in Fig 4, that shows the mean markers trajectories for an average individual (i.e random effects u_i set to 0). As expected, the level of the response is higher after a repeated number of injections for both binding antibody concentration and neutralization for all variants. Interestingly, the neutralization curves for BA.1, BA.2, and BA.5 are significantly

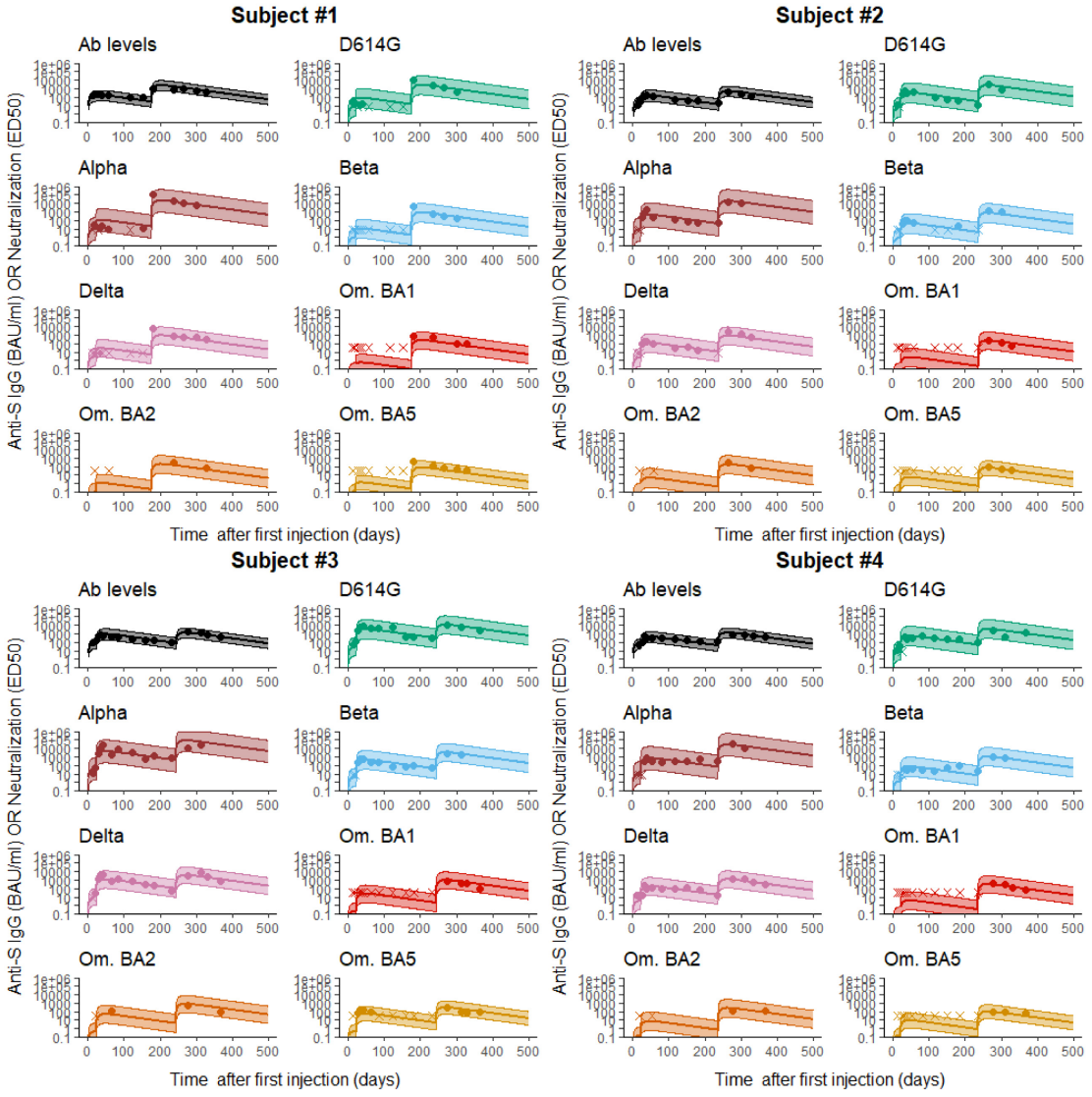


Fig 3. Individual fits for four representative individuals. The solid line is the subject-specific prediction and the shaded area is the 95% prediction interval. The plain dots and crosses represent the observed and censored data, respectively.

https://doi.org/10.1371/journal.pcbi.1011282.g003

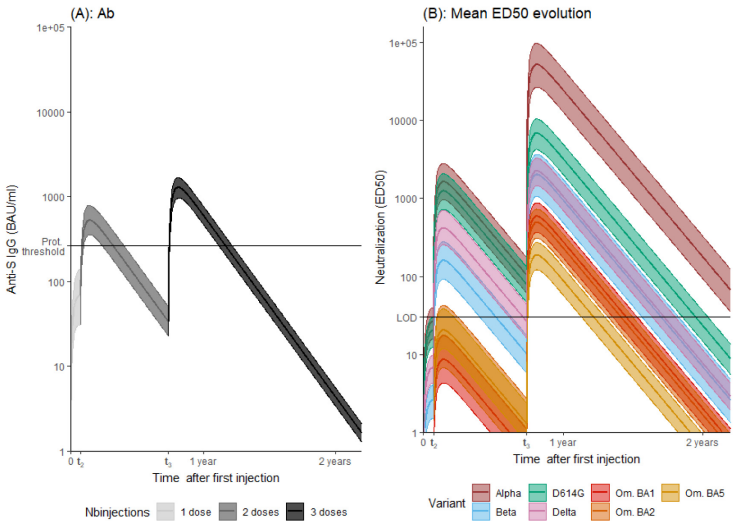


Fig 4. A: Predicted evolution of binding antibody concentration. The horizontal line corresponds to the value of 264 BAU/ml considered as a threshold against symptomatic infection. B: Predicted kinetics of ED^{v}_{50}. The horizontal line corresponds to the LOD. In all panels, the shaded area is the 95% prediction interval.

https://doi.org/10.1371/journal.pcbi.1011282.g004

lower than for the other variants, with no overlap in prediction intervals. The first and second doses elicit a neutralization response for Omicron (BA.1, BA.2, and BA.5) that remains below the detection limit in most individuals (which is consistent with the observed data) but is dramatically enhanced by the third injection. Regarding the ED^{v}_{50}/BAU ratio (See S2 Fig), we find that for the same concentration of binding antibodies, neutralization is significantly increased after each new injections for all variants and is significantly different for Alpha, D614G, {Delta, Beta} and {BA.1, BA.2, BA.5} variants.

Long-term predictions

As already shown in Fig 4, we can use the estimated models to predict the long-term trajectories of markers corresponding to the mean parameter values as well as 95% prediction intervals. It allows to derive an estimation of the time needed to reach a certain threshold after a three injections vaccination scheme with first vaccination at time $\mathbf{t_1} = 0$, second injection at time $\mathbf{t_2} = 27$ days and third injection at time $\mathbf{t_3} = 269$ days, corresponding to the mean observed time of injection in our cohort. Binding antibodies concentration is below 264 BAU/mL 154 (95/% PI [137; 173]) days after third vaccination. Neutralization reaches undetectable levels between 173 days (95/% PI [142; 200]) for BA.5 to 587 (95/% PI [537; 636]) for Alpha after the third dose.

Fig 5A displays the probability of having antibody concentration higher than the protection threshold established by Feng et al. [30] of 264 BAU/mL each days after the last injection in the counterfactual scenario where subjects only received one, two or three doses. The same is done for neutralizing activity again the VoCs (Fig 5B: one, Fig 5C: two or Fig 5D: three). It is possible to see the drastic effect of repeated injections on the levels reached by both binding

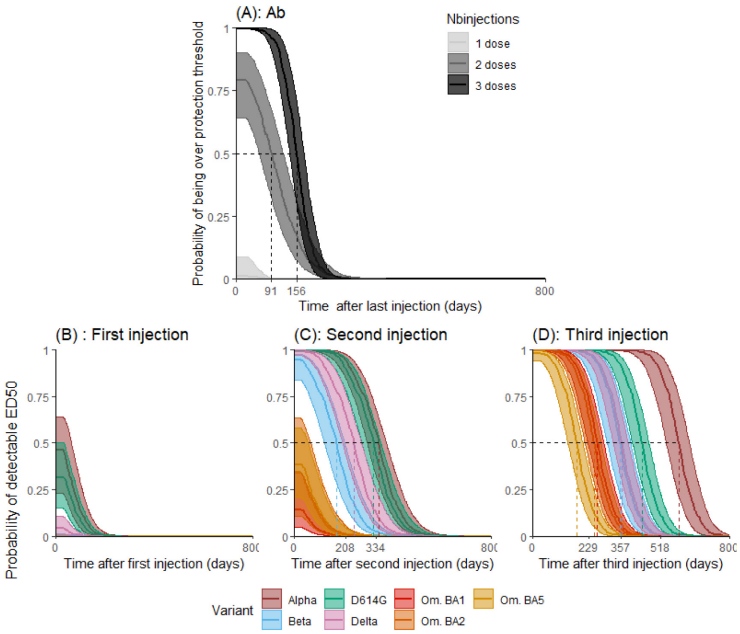


Fig 5. A: Predicted probability of having predicted antibody concentration (anti-S IgG) greater than 264 BAU/mL. B-C-D: Probability of having detectable neutralizing activity against VoCs after after the first (B), second (C) or third (D) vaccination dose. Simulations were performed assuming that the second and third vaccination doses occurred at day 27 and 269, respectively.

https://doi.org/10.1371/journal.pcbi.1011282.g005

antibodies concentration and neutralization for all variants. Strikingly, the full response duration is similar in length for the binding antibodies concentration after two or three injections. However, whereas response higher than 264 BAU/mL is reached in 100% (95% PI [99%; 100%]) of the population after three injections, it is only reached in 82% (95% PI [65%; 90%]) of the population after two doses, and never reached in the whole population after the first injection (value: 0% (95% PI [0%; 3%])). Table 4 provides the time needed for a proportion of a vaccinated population to return under a certain threshold. It explores multiple thresholds (100 BAU/mL, 264 BAU/mL, and 1000 BAU/mL for antibodies concentrations; and undetectability, 100 and 1000 for neutralization) that could be investigated when and if a clear level of correlate of protection is found. For all markers, there is a systematically and significantly higher duration of humoral activity after three compared to two injections. After three injections, duration of neutralization against Omicron variants (BA.1,BA.2 and BA.5) is significantly lower than for other variants for all thresholds.

Discussion

We proposed here a modeling framework to characterize the kinetics of antibodies to successive doses of Bnt162b2 vaccine. The originality of our approach is that we relied on both the kinetics of anti-S IgG binding antibodies and their neutralization against the major VoCs that

Table 4. Predicted distribution for the duration of anti-S IgG and neutralization activity above different threshold levels [95% prediction interval].

		Time to anti-S IgG		
	Population quantiles	<100 BAU/mL	<264 BAU/mL	<1000 BAU/mL
IgG	95%	163 [146; 185]	94 [74; 114]	0 [0; 0]
	50%	223 [207; 242]	152 [135; 173]	62 [32; 76]
	5%	289 [265; 305]	214 [193; 235]	119 [97; 138]
		Time to ED50		
Variant	Population quantiles	Undetectable	<100	<1000
D614G	95%	327 [294; 370]	252 [207; 282]	81 [39; 115]
	50%	433 [404; 471]	350 [316; 382]	186 [148; 217]
	5%	557 [506; 579]	460 [420; 491]	278 [252; 328]
Alpha	95%	476 [431; 528]	401 [341; 447]	230 [172; 283]
	50%	583 [539; 634]	499 [445; 551]	335 [280; 383]
	5%	706 [641; 745]	609 [550; 657]	427 [382; 491]
Beta	95%	237 [194; 290]	161 [108; 199]	0 [0; 0]
	50%	343 [301; 388]	259 [217; 299]	96 [44; 140]
	5%	467 [407; 499]	369 [323; 405]	188 [151; 247]
Delta	95%	245 [219; 289]	170 [132; 195]	0 [0; 0]
	50%	351 [323; 387]	267 [238; 297]	104 [67; 128]
	5%	475 [427; 497]	377 [342; 407]	196 [169; 241]
BA.1	95%	146 [117; 183]	70 [0; 97]	0 [0; 0]
	50%	252 [222; 287]	168 [139; 199]	0 [0; 0]
	5%	376 [324; 398]	278 [242; 309]	97 [69; 141]
BA.2	95%	132 [98; 181]	56 [0; 89]	0 [0; 0]
	50%	238 [203; 285]	155 [119; 191]	0 [0; 0]
	5%	362 [307; 392]	265 [223; 301]	83 [46; 131]
BA.5	95%	62 [0; 99]	0 [0; 0]	0 [0; 0]
	50%	168 [143; 203]	84 [54; 114]	0 [0; 0]
	5%	292 [246; 310]	194 [156; 225]	0 [0; 52]

https://doi.org/10.1371/journal.pcbi.1011282.t004

have emerged since 2021. Our model quantifies the benefit of successive injections and can be used to predict the duration of detectable neutralizing activity against each VoC. After the first dose, the model shows the significant action of each additional injection, especially of the third one, to increase the intrinsic antibody neutralizing quality against all VoCs [34]. However, both the maximum level achieved and the rate of decline could vary greatly between VoCs. Accordingly, the mean duration of detectable neutralizing activity after the third dose of vaccine was 20, 12, 8.5, 8 and 6 months for Alpha, Delta, and Omicron BA.1, BA.2 and BA.5 respectively. Our results also highlight the wide variability in patient response, with at least 5% of patients with undetectable neutralizing activity against Omicron BA.5 only 2 months after the third injection.

These results were obtained based on a number of hypotheses, which we summarize below. First, the model of antibody concentration dynamics remains simplified, with the memory compartment simply represented by a piecewise constant function over successive doses. In addition, the model assumes only one type of secreting cell population and thus overlooks the complexity of the B-cell response mechanism. Our model does not integrate the possible mechanisms causing the gain in neutralization observed over dose injections, and how this may modulated by the time between injections. For instance, it has been suggested that longer delay between injections could increase the Memory B-cells selection stringency in germinal

centers [36]. This could in turn lead to strategies to maximize antibody concentration, as suggested by theoretical models [37]. This is in line with measurements made on another cohorts, for example the one described in [38, 39] with a significantly longer delay between injections comparing to ours. In this case, measured antibody concentration after third injection was significantly higher than our prediction. Still, due to our data limitation (few subjects with similar vaccination schedule), we cannot isolate and thus estimate the effect of injection schedule on neutralizing activity. The choice of linear relationship between neutralizing activity and binding antibodies obviously omits features acting on neutralization which could have been incorporated in a more complete model, closer to biological mechanisms. Still, S3 Fig shows that this choice is in adequation with general trend in our data and testing more complex relationships, such as a sigmoid model, did not lead to statistical improvement (see Appendix B in S1 Appendix). Of note, this linear relationship also provided a good fit to the data on external cohorts [40–42]. Finally, the model assumed that the second dose would result in a similar change in protection for all variants. In the future, application of such approaches to larger populations of individuals, with a wider range of tested vaccination schedules, may allow some of these hypotheses to be relaxed and injection timing to be integrated into the model specification without compromising the identifiability of the parameters.

Due to the available data and constructed model, we restrict our analysis to the humoral part of the immune response triggered by vaccination only. Still, vaccination also induces a cellular immune response which may contribute to the clinical protection especially against the VoCs [43] (for a mechanistic model accounting for T-cell response, see Korosec et al. [19]). Regarding immune response induced by infection, to this date, 11 out of 26 followed subjects were infected with Omicron. This proportion is likely to increase as it is the case in the global vaccinated population. Thus, it is of great interest to model the hybrid protection induced by vaccination followed by natural infection. Still, due to model limitation, we discarded patient data after breakthrough infection. It would requires to deeply modify our model to integrate two different antibody populations, one coming from vaccination and targeting the historical strain and the other one targeting the Omicron variant. That is why this analysis is left to future works.

One of the main advantages of the model is its flexibility to easily incorporate information on new VoCs and to use the strength of information obtained on other viral variants to update the model as data become available. In fact, we continuously updated the model to include successive Omicron variants. Interestingly, despite the small number of samples available, a high degree of precision was achieved for all variants. For example, although patients had on average only two data points with detectable neutralization against the BA.2 variant, this was sufficient to achieve a good precision for the estimation of the model parameters (Table 2). Also, despite its simplicity, the estimated mechanistic model for antibody kinetics produces consistent predictions for patients with a different vaccination schedule than the one considered here (see Appendix D in S1 Appendix). Additionally, by using all available data (eg, by analyzing anti-S IgG and neutralizing activity of all patients simultaneously), the model reveals some signals of kinetics that were not visible when analyzing the individual markers separately. For example, we identified different slopes of antibody decline that directly affect the prediction of protection duration. Using the same data set and a simpler single-slope model, the time to undetectable neutralizing levels after the third dose of vaccine for D416G was estimated to be 11.5 months [21], which is shorter than our estimate of 13.5 months (derived from Fig 4). In their approach, Planas et al. [21] chose to adjust the anti-S IgG and ED_{50}^{v} decline separately for the different VoCs without considering causal relationships between them. On the contrary, our model assumes an influence of antibody concentration on the development of ED_{50}^{v}. Similar results are shown for the Delta variant (11.5 vs. 10.5 months, respectively), demonstrating

the importance of a model-based approach to predict neutralizing activity in the long term. Predictions for the Omicron strains were similar for BA.1 and BA.2, respectively 8.5 and 8 months, and show a reduction for BA.5 strains with 6 months. Interestingly, our results also suggest a longer duration of detectable neutralizing activity than what has been directly extrapolated from other observational studies [38, 44, 45], although this difference may also be due in part to the different experimental procedure used to measure neutralization.

The large and VoC-dependent variability in neutralization values argues for the use of individualized approaches to identify patients most at risk. Although it should be acknowledged that such an approach is hampered by the lack of an established neutralizing activity threshold as correlates of protection, its level was found to be associated with the risk of breakthrough infection. In a cohort of elderly nursing home residents, none of those the individuals with ED50 above 2136 had Omicron BA.1 breakthrough infection [46]. A model-based study found that a threshold of 1000 dramatically reduced peak viral load, suggesting that such a threshold may be a good indicator of protection against infection [40]. Interestingly, our results show that neutralizing levels for all Omicron variants remain largely below this value (Fig 5), consistent with the current understanding that BNT162b2 is poorly effective against disease acquisition in the Omicron era [39, 46]. Fortunately, the vaccine has shown high efficacy against severe disease to date [47, 48].

To date, the use of a fourth dose of vaccine to increase efficacy in France has been limited to high-risk patients who were not represented in this cohort. Nevertheless, we used the model to predict the neutralization levels that could be achieved after a fourth vaccine dose. Under the conservative assumption, yet consistent with available observational study [49], that this injection does not increase affinity or maturation parameters, our model predicts a similar duration of detectable neutralization as after the third dose, ranging from 172 to 256 days for the Omicron variants. Assuming that the fourth dose allows a similar increase in maturation and affinity as after the third dose, the model predicts that the duration of detectable neutralization could be much longer, ranging from 610 to 694 days for Omicron variants (see this supplementary analysis in Appendix E in S1 Appendix).

Supporting information

S1 Fig. ED_{50}^{v} raw data after one injection. Zoomed version of ED_{50}^{v} raw data presentation after one injection.
(EPS)

S2 Fig. Estimated mean evolution of $t \mapsto \frac{ED_{50}^{v}(t)}{Ab(t)}$. Evolution of ED_{50}^{v}:BAU ratio.
(EPS)

S3 Fig. Linear regressions $ED_{50}^{v} = \beta^{v} Ab$. Linear regressions $ED_{50}^{v} = \beta^{v} Ab$ for each VoC from simultaneously measured neutralizing activity and binding antibody concentration. The censored data have been removed.
(EPS)

S1 Appendix. Appendixes for "Modeling the evolution of the neutralizing antibody response against SARS-CoV-2 variants after several administrations of Bnt162b2".
(PDF)

S1 File. Available data. The dataset used for this analysis is available in the zip file Neutralization_Data_and_code.
(ZIP)

Acknowledgments

We thank Isabelle Staropoli, Florence Guivel-Benhassine, Françoise Porrot and all the members of the Virus and Immunity Unit for discussion and help, as well as Fabienne Peira, Vanessa Legros, Barbara De Dieuleveult, Aurelie Theillay, Sandra Pallay and Daniela Pires Roteia (CHR Orléans) for their help with the cohorts. Part of the experiments presented in this paper were carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d'Aquitaine (see https://www.plafrim.fr). We thank Simulations Plus, Lixoft division for the free academic use of the MonolixSuite.

Author Contributions

Conceptualization: Quentin Clairon, Mélanie Prague, Rodolphe Thiébaut, Jérémie Guedj.

Data curation: Delphine Planas, Timothée Bruel, Laurent Hocqueloux, Thierry Prazuck, Olivier Schwartz.

Formal analysis: Quentin Clairon.

Investigation: Quentin Clairon, Mélanie Prague, Delphine Planas, Timothée Bruel, Laurent Hocqueloux, Thierry Prazuck, Olivier Schwartz, Rodolphe Thiébaut, Jérémie Guedj.

Methodology: Quentin Clairon, Mélanie Prague, Rodolphe Thiébaut, Jérémie Guedj.

Resources: Delphine Planas, Timothée Bruel, Laurent Hocqueloux, Thierry Prazuck, Olivier Schwartz.

Software: Quentin Clairon.

Validation: Quentin Clairon, Mélanie Prague, Rodolphe Thiébaut, Jérémie Guedj.

Visualization: Quentin Clairon.

Writing – original draft: Quentin Clairon, Mélanie Prague, Olivier Schwartz, Jérémie Guedj.

Writing – review & editing: Quentin Clairon, Mélanie Prague, Delphine Planas, Timothée Bruel, Laurent Hocqueloux, Thierry Prazuck, Rodolphe Thiébaut, Jérémie Guedj.

References

1. Watson OJ, Barnsley G, Toor J, Hogan AB, Winskill P, Ghani AC. Global impact of the first year of COVID-19 vaccination: a mathematical modelling study. The Lancet Infectious Diseases. 2022; 22 (9):1293–1302. https://doi.org/10.1016/S1473-3099(22)00320-6 PMID: 35753318
2. Thomas SJ, Moreira ED Jr, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine through 6 months. New England Journal of Medicine. 2021; 385(19):1761–1773. https://doi.org/10.1056/NEJMoa2110345 PMID: 34525277
3. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. New England journal of medicine. 2020;. https://doi.org/10.1056/NEJMoa2034577 PMID: 33301246
4. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. New England journal of medicine. 2020;. https://doi.org/10.1056/NEJMoa2035389 PMID: 33378609
5. Harris RJ, Hall JA, Zaidi A, Andrews NJ, Dunbar JK, Dabrera G. Effect of vaccination on household transmission of SARS-CoV-2 in England. New England Journal of Medicine. 2021; 385(8):759–760. https://doi.org/10.1056/NEJMc2107717 PMID: 34161702
6. Pritchard E, Matthews PC, Stoesser N, Eyre DW, Gethings O, Vihta KD, et al. Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. Nature medicine. 2021; 27(8):1370–1378. https://doi.org/10.1038/s41591-021-01410-w PMID: 34108716

7. Eyre DW, Taylor D, Purver M, Chapman D, Fowler T, Pouwels KB, et al. Effect of Covid-19 vaccination on transmission of alpha and delta variants. New England Journal of Medicine. 2022; 386(8):744–756. https://doi.org/10.1056/NEJMoa2116597 PMID: 34986294
8. Gupta S, Cantor J, Simon KI, Bento AI, Wing C, Whaley CM. Vaccinations Against COVID-19 May Have Averted Up To 140,000 Deaths In The United States: Study examines role of COVID-19 vaccines and deaths averted in the United States. Health Affairs. 2021; 40(9):1465–1472.
9. Meslé MM, Brown J, Mook P, Hagan J, Pastore R, Bundle N, et al. Estimated number of deaths directly averted in people 60 years and older as a result of COVID-19 vaccination in the WHO European Region, December 2020 to November 2021. Eurosurveillance. 2021; 26(47):2101021. https://doi.org/10.2807/1560-7917.ES.2021.26.47.2101021 PMID: 34823641
10. Cele S, Jackson L, Khoury DS, Khan K, Moyo-Gwete T, Tegally H, et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. Nature. 2022; 602(7898):654–656. https://doi.org/10.1038/s41586-021-04387-1 PMID: 35016196
11. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature. 2021; 596(7871):276–280. https://doi.org/10.1038/s41586-021-03777-9 PMID: 34237773
12. Bernal JL, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, et al. Effectiveness of Covid-19 vaccines against the B. 1.617. 2 (Delta) variant. New England Journal of Medicine. 2021;.
13. Andrews N, Stowe J, Kirsebom F, Toffa S, Rickeard T, Gallagher E, et al. Covid-19 vaccine effectiveness against the Omicron (B. 1.1. 529) variant. New England Journal of Medicine. 2022; 386(16):1532–1546. https://doi.org/10.1056/NEJMoa2119451 PMID: 35249272
14. Mizrahi B, Lotan R, Kalkstein N, Peretz A, Perez G, Ben-Tov A, et al. Correlation of SARS-CoV-2-breakthrough infections to time-from-vaccine. Nature communications. 2021; 12(1):1–5. https://doi.org/10.1038/s41467-021-26672-3 PMID: 34737312
15. Goldberg Y, Mandel M, Bar-On YM, Bodenheimer O, Freedman L, Haas EJ, et al. Waning immunity after the BNT162b2 vaccine in Israel. New England Journal of Medicine. 2021; 385(24):e85. https://doi.org/10.1056/NEJMoa2114228 PMID: 34706170
16. Levin EG, Lustig Y, Cohen C, Fluss R, Indenbaum V, Amit S, et al. Waning immune humoral response to BNT162b2 Covid-19 vaccine over 6 months. New England Journal of Medicine. 2021; 385(24):e84. https://doi.org/10.1056/NEJMoa2114583 PMID: 34614326
17. Cromer D, Juno JA, Khoury D, Reynaldi A, Wheatley AK, Kent SJ, et al. Prospects for durable immune control of SARS-CoV-2 and prevention of reinfection. Nature Reviews Immunology. 2021; 21(6):395–404. https://doi.org/10.1038/s41577-021-00550-x PMID: 33927374
18. Zhu F, Althaus T, Tan CW, Costantini A, Chia WN, Chau NVV, et al. WHO international standard for SARS-CoV-2 antibodies to determine markers of protection. The Lancet Microbe. 2022; 3(2):e81–e82. https://doi.org/10.1016/S2666-5247(21)00307-4 PMID: 34901897
19. Korosec CS, Farhang-Sardroodi S, Dick DW, Gholami S, Ghaemi MS, Moyles IR, et al. Long-term durability of immune responses to the BNT162b2 and mRNA-1273 vaccines based on dosage, age and sex. Scientific Reports. 2022; 12(1):21232. https://doi.org/10.1038/s41598-022-25134-0 PMID: 36481777
20. Planas D, Bruel T, Grzelak L, Guivel-Benhassine F, Staropoli I, Porrot F, et al. Sensitivity of infectious SARS-CoV-2 B. 1.1. 7 and B. 1.351 variants to neutralizing antibodies. Nature medicine. 2021; 27(5):917–924. https://doi.org/10.1038/s41591-021-01318-5 PMID: 33772244
21. Planas D, Staropoli I, Porot F, Guivel-Benhassine F, Handala L, Prot M, et al. Duration of BA. 5 neutralization in sera and nasal swabs from SARS-CoV-2 vaccinated individuals, with or without Omicron breakthrough infection. Med. 2022;. https://doi.org/10.1016/j.medj.2022.09.010 PMID: 36228619
22. Pasin C, Balleli I, Effelterre TV, Bockstal V, Solforosi L, Prague M, et al. Dynamic of the humoral immune response to a prime-boost Ebola vaccine: quantification and source of variation. Journal of Virology. 2019; 93(18):e00579–19. https://doi.org/10.1128/JVI.00579-19 PMID: 31243126
23. Balelli I, Pasin C, Prague M, Crauste F, Thiébaut R. A model for establishment, maintenance and reactivation of the immune response after two-dose vaccination regimens against Ebola virus. Journal of Theoretical Biology. 2020; p. 110254.
24. Muecksch F, Wang Z, Cho A, Gaebler C, Tanfous TB, DaSilva J, et al. Increased potency and breadth of SARS-CoV-2 neutralizing antibodies after a third mRNA vaccine dose. bioRxiv. 2022;. https://doi.org/10.1101/2022.02.14.480394 PMID: 35194607
25. Wang K, Jia Z, Bao L, Wang L, Cao L, Chi H, et al. Memory B cell repertoire from triple vaccinees against diverse SARS-CoV-2 variants. Nature. 2022; 603(7903):919–925. https://doi.org/10.1038/s41586-022-04466-x PMID: 35090164

26. Padmanabhan P, Desikan R, Dixit NM. Modeling how antibody responses may determine the efficacy of COVID-19 vaccines. Nature Computational Science. 2022; 2(2):123–131. https://doi.org/10.1038/s43588-022-00198-0

27. Murphy SA, Van der Vaart AW. On profile likelihood. Journal of the American Statistical Association. 2000; 95(450):449–465. https://doi.org/10.2307/2669392

28. Lixoft. Monolix version 2020R1. Antony, France. 2020;.

29. Samson A, Lavielle M, Mentré F. Extension of the SAEM algorithm to left-censored data in nonlinear mixed-effects model: Application to HIV dynamics model. Computational Statistics & Data Analysis. 2006; 51(3):1562–1574. https://doi.org/10.1016/j.csda.2006.05.007

30. Feng S, Phillips DJ, White T, Sayal H, Aley PK, Bibi S, et al. Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. Nature medicine. 2021; 27(11):2032–2040. https://doi.org/10.1038/s41591-021-01540-1 PMID: 34588689

31. Gilbert PB, Montefiori DC, McDermott AB, Fong Y, Benkeser D, Deng W, et al. Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. Science. 2022; 375(6576):43–50. https://doi.org/10.1126/science.abm3425 PMID: 34812653

32. Alexandre M, Marlin R, Prague M, Coleon S, Kahlaoui N, Cardinaud S, et al. Modelling the response to vaccine in non-human primates to define SARS-CoV-2 mechanistic correlates of protection. Elife. 2022; 11:e75427. https://doi.org/10.7554/eLife.75427 PMID: 35801637

33. Henningsen A. Estimating censored regression models in R using the censReg Package. R package vignettes. 2010; 5:12.

34. Goel RR, Apostolidis SA, Painter MM, Mathew D, Pattekar A, Kuthuru O, et al. Distinct antibody and memory B cell responses in SARS-CoV-2 naïve and recovered individuals after mRNA vaccination. Science immunology. 2021; 6(58):eabi6950.

35. Booth BJ, Ramakrishnan B, Narayan K, Wollacott AM, Babcock GJ, Shriver Z, et al. Extending human IgG half-life using structure-guided design. In: MAbs. vol. 10. Taylor & Francis; 2018. p. 1098–1110.

36. Garg AK, Mittal S, Padmanabhan P, Desikan R, Dixit NM. Increased B cell selection stringency in germinal centers can explain improved COVID-19 vaccine efficacies with low dose prime or delayed boost. Frontiers in immunology. 2021; p. 5064. https://doi.org/10.3389/fimmu.2021.776933 PMID: 34917089

37. Stolfi P, Castiglione F, Mastrostefano E, Di Biase I, Di Biase S, Palmieri G, et al. In-silico evaluation of adenoviral COVID-19 vaccination protocols: Assessment of immunological memory up to 6 months after the third dose. Frontiers in Immunology. 2022; 13. https://doi.org/10.3389/fimmu.2022.998262 PMID: 36353634

38. Brockman MA, Mwimanzi F, Lapointe HR, Sang Y, Agafitei O, Cheung PK, et al. Reduced magnitude and durability of humoral immune responses to COVID-19 mRNA vaccines among older adults. The Journal of Infectious Diseases. 2022; 225(7):1129–1140. https://doi.org/10.1093/infdis/jiab592 PMID: 34888688

39. Lapointe HR, Mwimanzi F, Cheung PK, Sang Y, Yaseen F, Kalikawe R, et al. Serial infection with SARS-CoV-2 Omicron BA. 1 and BA. 2 following three-dose COVID-19 vaccination. medRxiv. 2022;. https://doi.org/10.3389/fimmu.2022.947021 PMID: 36148225

40. Lingas G, Planas D, Péré H, Duffy D, Staropoli I, Porrot F, et al. Modelling the association between neutralizing antibody levels and SARS-CoV-2 viral dynamics: implications to define correlates of protection against infection. medRxiv. 2023; p. 2023–03.

41. Tran TT, Vaage EB, Mehta A, Chopra A, Tietze L, Kolderup A, Anthi A, König M, Nygaard G, Lind A, et al. Titers of antibodies against ancestral SARS-CoV-2 correlate with levels of neutralizing antibodies to multiple variants. npj Vaccines. 2022; 7(1):174 https://doi.org/10.1038/s41541-022-00586-7 PMID: 36585405

42. Manenti A, Gianchecchi E, Dapporto F, Leonardi M, Cantaloni P, Fattorini F, Piu P, Bollati V, Pastorino U, Apolone G, et al. Evaluation and correlation between SARS-CoV-2 neutralizing and binding antibodies in convalescent and vaccinated subjects. Journal of Immunological Methods. 2022; 500:113197 https://doi.org/10.1016/j.jim.2021.113197 PMID: 34843712

43. Li C, Lee A, Grigoryan L, Arunachalam PS, Scott MK, Trisal M, et al. Mechanisms of innate and adaptive immunity to the Pfizer-BioNTech BNT162b2 vaccine. Nature Immunology. 2022; 23(4):543–555. https://doi.org/10.1038/s41590-022-01163-9 PMID: 35288714

44. Mwimanzi F, Lapointe HR, Cheung PK, Sang Y, Yaseen F, Umviligihozo G, et al. Older adults mount less durable humoral responses to two doses of COVID-19 mRNA vaccine, but strong initial responses to a third dose. medRxiv. 2022;.

45. Mwimanzi FM, Lapointe HR, Cheung PK, Sang Y, Yaseen F, Kalikawe R, et al. Brief Report: Impact of age and SARS-CoV-2 breakthrough infection on humoral immune responses after three doses of COVID-19 mRNA vaccine. medRxiv. 2022;.

46. Bruel T, Pinaud L, Tondeur L, Planas D, Staropoli I, Porrot F, et al. Neutralising antibody responses to SARS-CoV-2 omicron among elderly nursing home residents following a booster dose of BNT162b2 vaccine: A community-based, prospective, longitudinal cohort study. EClinicalMedicine. 2022; 51:101576. https://doi.org/10.1016/j.eclinm.2022.101576 PMID: 35891947

47. Hacisuleyman E, Hale C, Saito Y, Blachere NE, Bergh M, Conlon EG, et al. Vaccine breakthrough infections with SARS-CoV-2 variants. New England Journal of Medicine. 2021; 384(23):2212–2218. https://doi.org/10.1056/NEJMoa2105000 PMID: 33882219

48. Kuhlmann C, Mayer CK, Claassen M, Maponga T, Burgers WA, Keeton R, et al. Breakthrough infections with SARS-CoV-2 omicron despite mRNA vaccine booster dose. The Lancet. 2022; 399 (10325):625–626. https://doi.org/10.1016/S0140-6736(22)00090-3 PMID: 35063123

49. Regev-Yochay G, Gonen T, Gilboa M, Mandelboim M, Indenbaum V, Amit S, et al. Efficacy of a fourth dose of COVID-19 mRNA vaccine against omicron. New England Journal of Medicine. 2022; 386 (14):1377–1380. https://doi.org/10.1056/NEJMc2202542 PMID: 35297591

6.6 Alexandre et al. 2022 (eLife) Correlate of Protection against SARS-CoV-2

Modeling the response to vaccine in non-human primates to define SARS-CoV-2 mechanistic correlates of protection. Alexandre M., Marlin R.*, **Prague M.***, Coléon S.,... and Thiébaut R. *eLife.* 11 - e7542 - July 20222.

This article integrates in the second axis of my research "Within-host Modeling in Infectious Diseases". It has been written by a PhD student that I co-directed.

I selected this paper because it is a good example of mechanistic modeling of virus dynamics. Moreover, it demonstrates the use of mechanistic models to causally identify a correlate of protection against an infection from various markers. A significant challenge was integrating Prentice's criteria for defining a surrogate marker within the mechanistic model framework. This integration led to the conclusion that the neutralization capacities of antibodies are crucial (i.e. functional feature as opposed to quantification of antibody concentrations), corroborating findings from others on an original data in NHP studies on a new vaccine platform, with validation is external studies.

RESEARCH ARTICLE

Modelling the response to vaccine in non-human primates to define SARS-CoV-2 mechanistic correlates of protection

Marie Alexandre[1], Romain Marlin[2†], Mélanie Prague[1†], Severin Coleon[3,4], Nidhal Kahlaoui[2], Sylvain Cardinaud[3,4], Thibaut Naninck[2], Benoit Delache[2], Mathieu Surenaud[3,4], Mathilde Galhaut[2], Nathalie Dereuddre-Bosquet[2], Mariangela Cavarelli[2], Pauline Maisonnasse[2], Mireille Centlivre[3,4], Christine Lacabaratz[3,4], Aurelie Wiedemann[3,4], Sandra Zurawski[5], Gerard Zurawski[5], Olivier Schwartz[3,6,7], Rogier W Sanders[8], Roger Le Grand[2], Yves Levy[3,4,9], Rodolphe Thiébaut[1,3,10]*

[1]University of Bordeaux, Department of Public Health, Inserm Bordeaux Population Health Research Centre, Inria SISTM, Bordeaux, France; [2]Center for Immunology of Viral, Auto-immune, Hematological and Bacterial Diseases (IMVA-HB/IDMIT), Université Paris-Saclay, Inserm, CEA, Fontenay-aux-Roses, France; [3]Vaccine Research Institute, Créteil, France; [4]Inserm U955, Créteil, France; [5]Baylor Scott and White Research Institute, Dallas, United States; [6]Virus & Immunity Unit, Department of Virology, Institut Pasteur, Paris, France; [7]CNRS UMR 3569, Paris, France; [8]Department of Medical Microbiology, Amsterdam UMC, University of Amsterdam Amsterdam Infection & Immunity Institute, Amsterdam, Netherlands; [9]AP-HP, Hôpital Henri-Mondor Albert-Chenevier, Service d'Immunologie Clinique et Maladies Infectieuses, Créteil, France; [10]CHU Bordeaux, Department of Medical information, Bordeaux, France

***For correspondence:** rodolphe.thiebaut@u-bordeaux.fr

[†]These authors contributed equally to this work

Competing interest: The authors declare that no competing interests exist.

Funding: See page 20

Preprinted: 01 November 2021
Received: 09 November 2021
Accepted: 22 June 2022
Published: 08 July 2022

Reviewing Editor: Frederik Graw, Heidelberg University, Germany

Abstract The definition of correlates of protection is critical for the development of next-generation SARS-CoV-2 vaccine platforms. Here, we propose a model-based approach for identifying mechanistic correlates of protection based on mathematical modelling of viral dynamics and data mining of immunological markers. The application to three different studies in non-human primates evaluating SARS-CoV-2 vaccines based on CD40-targeting, two-component spike nanoparticle and mRNA 1273 identifies and quantifies two main mechanisms that are a decrease of rate of cell infection and an increase in clearance of infected cells. Inhibition of RBD binding to ACE2 appears to be a robust mechanistic correlate of protection across the three vaccine platforms although not capturing the whole biological vaccine effect. The model shows that RBD/ACE2 binding inhibition represents a strong mechanism of protection which required significant reduction in blocking potency to effectively compromise the control of viral replication.

Editor's evaluation

This work should be of interest to a broad readership in infectious diseases, especially those people interested in modeling of infections. It combines statistical and mechanistic modeling to find assayable correlates of immunity for vaccines. This method could be relevant to many diseases or vaccines, although the particular markers identified here likely will be limited in their generalizability.

Introduction

There is an unprecedented effort for SARS-CoV-2 vaccine development with 294 candidates currently evaluated (***World Health Organization, 2021***). However, variants of concern have emerged before the vaccine coverage was large enough to control the pandemics (***Cobey et al., 2021***). Despite a high rate of vaccine protection, these variants might compromise the efficacy of current vaccines (***Kuzmina et al., 2021***; ***Planas et al., 2021***; ***Lustig et al., 2021***; ***Zhou et al., 2021***). Control of the epidemic by mass vaccination may also be compromised by unknown factors such as long-term protection and the need of booster injections in fragile, immuno-compromised, elderly populations, or even for any individual if protective antibody levels wane. Furthermore, the repeated use of some of the currently approved vaccine could be compromised by potential adverse events or by immunity against vaccine viral vectors (***Greinacher et al., 2021***). Finally, the necessity to produce the billions of doses required to vaccinate the world's population also explains the need to develop additional vaccine candidates.

The identification of correlates of protection (CoPs) is essential to accelerate the development of new vaccines and vaccination strategies (***Koch et al., 2021***; ***Jin et al., 2021***). Binding antibodies to SARS-CoV-2 and in vitro neutralization of virus infection are clearly associated with protection (***Khoury et al., 2021***; ***Yu et al., 2020***; ***Earle et al., 2021***; ***Feng et al., 2021***). However, the respective contribution to virus control in vivo remains unclear (***Zost et al., 2020***), and many other immunological mechanisms may also be involved, including other antibody-mediated functions (antibody-dependent cellular cytotoxicity [ADCC], antibody-dependent complement deposition [ADCD], antibody-dependent cellular phagocytosis [ADCP]; ***Yu et al., 2020***; ***Mercado et al., 2020***; ***Tauzin et al., 2021***), as well as T cell immunity (***McMahan et al., 2021***). Furthermore, CoP may vary between the vaccine platforms (***Plotkin, 2013***; ***Plotkin, 2020***; ***Bradfute and Bavari, 2011***; ***Dagotto et al., 2020***).

Non-human primate (NHP) studies offer a unique opportunity to evaluate early markers of protective response (***Muñoz-Fontela, 2020***; ***Eyal and Lipsitch, 2021***). Challenge studies in NHP allow the evaluation of vaccine impact on the viral dynamics in different tissue compartments (upper and lower respiratory tract) from day 1 of virus exposure (***Yu et al., 2020***; ***Mercado et al., 2020***; ***Corbett et al., 2020***). Such approaches in animal models may thus help to infer, for example, the relation between early viral events and disease or the capacity to control secondary transmissions.

Here, we propose to apply a model-based approach on NHP studies to evaluate (i) the immune mechanisms involved in the vaccine response and (ii) the markers capturing this/these effect(s) leading to identification of mechanisms of protection and definition of mechanistic CoP (***Plotkin and Gilbert, 2012***). First, we present a mechanistic approach based on ordinary differential equation (ODE) models reflecting the virus-host interaction inspired from models proposed for SARS-CoV-2 infection (***Gonçalves et al., 2020***; ***Kim et al., 2021***; ***Gonçalves et al., 2021***; ***Wang et al., 2020***; ***Marc et al., 2021***; ***Ke et al., 2021***) and other viruses (***Myers et al., 2021***; ***Baccam et al., 2006***; ***Goyal et al., 2019***; ***Goyal et al., 2017***). The proposed model includes several new aspects refining the modelling of viral dynamics in vivo, in addition to the integration of vaccine effect. A specific inoculum compartment allows distinguishing the virus coming from the challenge inoculum and the virus produced de novo, which is a key point in the context of efficacy provided by antigen-specific pre-existing immune effectors induced by the vaccine. Then, an original data mining approach is implemented to identify the immunological biomarkers associated with specific mechanisms of vaccine-induced protection.

We apply our approach to a recently published study (***Marlin et al., 2021***) testing a protein-based vaccine targeting the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein to CD40 (αCD40.RBD vaccine). Targeting vaccine antigens to dendritic cells via the surface receptor CD40 represents an appealing strategy to improve subunit-vaccine efficacy (***Flamar et al., 2012***; ***Zurawski et al., 2017***; ***Cheng et al., 2018***; ***Godot et al., 2020***) and for boosting natural immunity in SARS-CoV-2 convalescent NHP.

We show that immunity induced by natural SARS-CoV-2 infection, as well as vaccine-elicited immune responses contribute to viral load control by (i) blocking new infection of target cells and (ii) by increasing the loss of infected cells. The modelling showed that antibodies inhibiting binding of RBD to ACE2 correlated with blockade of new infections and RBD-binding antibodies correlate with the loss of infected cells, reflecting importance of additional antibody functionalities. The role of RBD/ACE2-binding inhibition has been confirmed in two other vaccine platforms.

Results

A new mechanistic model fits the in vivo viral load dynamics in nasopharyngeal and tracheal compartments

The mechanistic model aims at capturing the viral dynamics following challenge with SARS-CoV-2 virus in NHP. For that purpose, we used data obtained from 18 cynomolgus macaques involved in the vaccine study reported by ***Marlin et al., 2021***, and exposed to a high dose (1×10^6 pfu) of SARS-CoV-2 administered via the combined intra-nasal and intra-tracheal route. The viral dynamics during the primary infections were characterized by a peak of genomic RNA (gRNA) production 3 days post-infection in both tracheal and nasopharyngeal compartments, followed by a decrease toward undetectable levels beyond day 15 (***Figure 1—figure supplement 1***). At the convalescent phase (median 24 weeks after the primary infection), 12 macaques were challenged with SARS-CoV-2 a second time, 4 weeks after being randomly selected to receive either a placebo (n=6) or a single injection of the αCD40.RBD vaccine (n=6) (***Figure 1A***). A third group of six naïve animals were infected at the same time. Compared to this naïve group, viral dynamics were blunted following the second challenge of convalescent animals with the lowest viral load observed in vaccinated animals (***Figure 1B***, ***Figure 1—figure supplement 2***).

We developed a mathematical model to better characterize the impact of the immune response on the viral gRNA and subgenomic RNA (sgRNA) dynamics, adapted from previously published work (***Gonçalves et al., 2020***; ***Kim et al., 2021***; ***Baccam et al., 2006***), which includes uninfected target cells (T) that can be infected (I_1) and produce virus after an eclipse phase (I_2). The virus generated can be infectious (V_i) or non-infectious (V_{ni}). Although a single compartment for de novo produced viruses (V) could be mathematically considered, two distinct ODE compartments were assumed for a better understanding of the model. We completed the model by a compartment for the inoculum to distinguish between the injected virus (V_s) and the virus produced de novo by the host (V_i and V_{ni}). In both compartments of the upper respiratory tract (URT), the trachea and nasopharynx, viral dynamics were distinctively described by this model (***Figure 2A***). Viral exchange between the two compartments was tested (either from the nasopharynx to the trachea or vice versa). However, as described in the literature (***Gonçalves et al., 2021***; ***Ke et al., 2020***; ***Pinky et al., 2021***) and demonstrated by the additional modelling work in Appendix 1 'Model building', viral transport within the respiratory tract plays a negligible role in viral kinetics compared with viral clearance. Consequently, no exchange was considered in the model. Using the gRNA and sgRNA viral loads, we jointly estimated (i.e., shared random effects and covariates) the viral infectivity (β), the viral production rate (P), and the loss rate of infected cells (δ) in the two compartments. We assumed that gRNA and sgRNA were proportional to the free virus and the infected cells, respectively. This modelling choice relied on both biological and mathematical reasons (see section Materials and methods for more details). Due to identifiability issues, the duration of the eclipse phase ($1/k$), the clearance of free viruses from the inoculum (c_i) and produced de novo (c) were estimated separately by profile likelihood and assumed to be identical in the two compartments of the URT. In addition, infectious and non-infectious viruses were assumed to be cleared at the same rate. We estimated the viral infectivity at 0.95×10^{-6} ($CI_{95\%}$ [0.18×10^{-6}; 4.94×10^{-6}]) $(copies/mL)^{-1}$ day^{-1} in naïve animals, which is in the range of previously reported modelling results whether in the case of SARS-CoV-2 virus (***Kim et al., 2021***; ***Wang et al., 2020***) or influenza (***Myers et al., 2021***; ***Baccam et al., 2006***). We found estimates of the loss rates of infected cells of 1.04 ($CI_{95\%}$ [0.79; 1.37]) day^{-1}, corresponding to a mean half-life of 0.67 day. This estimation was consistent with previously published results obtained on SARS-CoV-2 virus showing the mean value of this parameter ranging from 0.60 to 2 day^{-1} (i.e., half-life between 0.35 and 1.16 days) (***Gonçalves et al., 2020***; ***Kim et al., 2021***; ***Gonçalves et al., 2021***; ***Wang et al., 2020***; ***Marc et al., 2021***). The eclipse phase (3 day^{-1}) was found similar to the values commonly used in the literature (***Gonçalves et al., 2020***; ***Marc et al., 2021***; ***Myers et al., 2021***; ***Baccam et al., 2006***). Here, we distinguished the clearance of the inoculum which was much higher (20 virions day^{-1}) as compared to the clearance of the virus produced de novo (3 virions day^{-1}). While the half-life of the virus de novo produced usually approximates 1.7 hr (i.e., c=10 day^{-1}) (***Gonçalves et al., 2020***; ***Gonçalves et al., 2021***; ***Marc et al., 2021***; ***Myers et al., 2021***), because of this distinction, our model provided a higher estimation of 5.5 hr which remained in accordance with the estimations obtained by ***Baccam et al., 2006***, on influenza A. Furthermore, the viral production by each infected cells was estimated to be higher in the nasopharyngeal compartment (12.1×10^3 virions $cell^{-1}$ day^{-1}, $CI_{95\%}$ [3.15×10^3; 46.5×10^3]) as

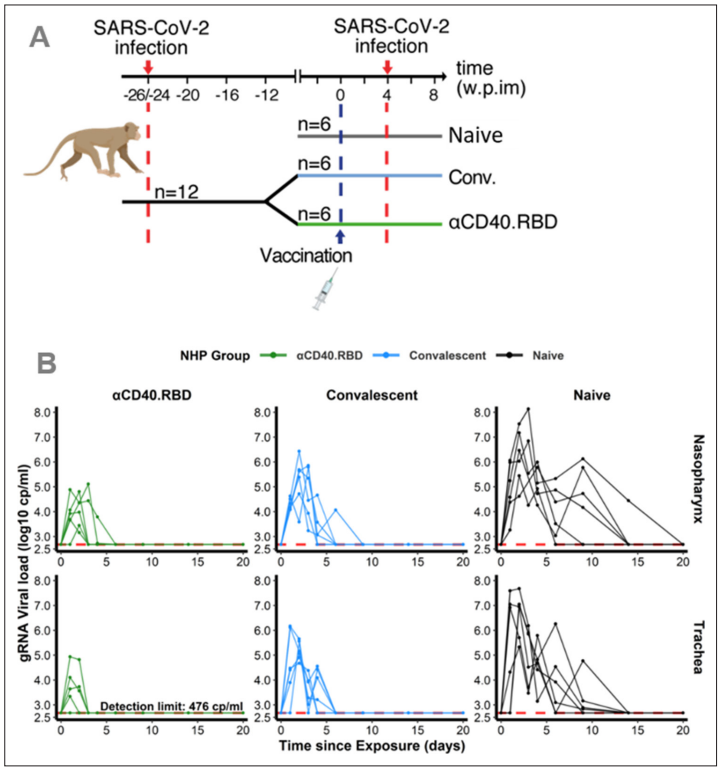


Figure 1. Design of the study 1 and viral dynamics. (**A**) *Study design*. Cynomolgus macaques (*Macaca fascicularis*), aged 37–58 months (8 females and 13 males). 24–26 weeks post-infection with SARS-CoV-2, 12 of these animals were randomly assigned in two experimental groups. The convalescent-vaccinated group (*n*=6) received 200 µg of αCD40.RBD vaccine. The other six convalescent animals were used as controls. Additional six age matched (43.7 months±6.76) cynomolgus macaques from same origin were included in the study as controls naïve from any exposure to SARS-CoV-2. Four weeks after immunization, all animals were exposed to a total dose of 10^6 pfu of SARS-CoV-2 virus via the combination of intra-nasal and intra-tracheal routes. In this work, only data collected from the second exposure were considered. (**B**) Individual $\log_{10}$ transformed genomic RNA (gRNA) viral load dynamics in nasopharyngeal swabs (top) and tracheal swabs (bottom) after the initial exposure to SARS-CoV-2 in naïve macaques (black, right) and after the second exposure in convalescent (blue, middle) and αCD40.RBD-vaccinated convalescent (green, left) groups. Horizontal red dashed lines indicate the limit of quantification.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1. Genomic RNA (gRNA) viral load longitudinally measured in the trachea and nasopharynx after the second exposure in the study 1.

Source data 2. Genomic RNA (gRNA) viral load longitudinally measured in the trachea and nasopharynx after the first exposure for convalescent non-human primates (NHPs) in the study 1.

Source data 3. Anti-spike IgG longitudinally measured post-immunization and quantified by Luminex in the study 1.

Source data 4. Quantification of the spike/ACE2-binding inhibition longitudinally measured post-immunization and quantified by Mesoscale Discovery (MSD) assay (in 1/ECL) in the study 1.

Source data 5. Anti-N and anti-RBD binding antibodies longitudinally measured post-immunization and quantified by Mesoscale Discovery (MSD) assay (in AU mL^{-1}) in the study 1.

Source data 6. Subgenomic RNA (sgRNA) viral load longitudinally measured in the trachea and nasopharynx after

Figure 1 continued on next page

Figure 1 continued

the second exposure in the study 1.

Source data 7. Antigen-specific T-cell response longitudinally measured post-exposure in % of CD4+ T cells measured by ICS in the study 1.

Source data 8. Antigen-specific T-cell response longitudinally measured post-exposure in % of CD8+ T cells measured by ICS in the study 1.

Source data 9. T-cell response expressing IFN-γ longitudinally measured post-exposure by ELISpot in the study 1.

Source data 10. Cytokine concentrations measured post-exposure in the study 1.

Source data 11. Quantification of the neutralization function of antibodies against three variants (B117, B1351, and D614G) longitudinally measured post-exposition (in ED50) in the study 1.

Figure supplement 1. Viral dynamics after the first exposure to SARS-CoV-2 and biomarker measurements from the first to the second exposure to SARS-CoV-2.

Figure supplement 2. Subgenomic viral dynamics after the second exposure to SARS-CoV-2.

Figure supplement 3. Antibody measurements after the second exposure to SARS-CoV-2.

Figure supplement 4. Antigen-specific T-cell responses in non-human primates (NHPs) after the second exposure to SARS-CoV-2.

Figure supplement 5. Cytokines and chemokines in the plasma in non-human primates (NHPs) after the second exposure to SARS-CoV-2.

compared to the tracheal compartment (0.92×10^3 virions cell^{-1} day^{-1}, $CI_{95\%}$ [0.39×10^3; 2.13×10^3]). These estimations are in agreement with the observation of the intense production of viral particles by primary human bronchial epithelial cells in culture (***Robinot et al., 2021***). In particular, they are in the range of estimates obtained within the URT, either in NHP (***Gonçalves et al., 2021***) or in humans (***Wang et al., 2020***), with the product $p \times T_0$ equals to 15.1×10^8 ($CI_{95\%}$ [3.98×10^8; 58.1×10^8]) and 0.21×10^8 ($CI_{95\%}$ [0.088×10^8; 0.48×10^8]) virions mL^{-1} day^{-1} in the nasopharynx and the trachea, respectively. By allowing parameters to differ between animals (through random effects), the variation of cell infectivity and of the loss rate of infected cells captured the observed variation of the dynamics of viral load. The variation of those parameters could be partly explained by the group to which the animals belong reducing the unexplained variability of the cell infectivity by 66% and of the loss rate of infected cells by 54% (***Supplementary file 1***). The model fitted well the observed dynamics of gRNA and sgRNA (***Figure 2B***).

Modelling of the dynamics of viral replication argues for the capacity of αCD40.RBD vaccine to block virus entry into host cells and to promote the destruction of infected cells

We distinguish the respective contribution of the vaccine effect and post-infection immunity on the reduction of the cell infection rate and the increase of the clearance of infected cells. Because blocking de novo infection and promoting the destruction of infected cells would lead to different viral dynamics profile (***Figure 2—figure supplement 1***), we were able to identify the contribution of each mechanism by estimating the influence of the vaccine compared to placebo or naïve animals on each model parameter. The αCD40.RBD vaccine reduced by 99.6% the infection of target cells in the trachea compared to the naïve group. The estimated clearance of infected cells was 1.04 day^{-1} (95% CI 0.75; 1.45) in naïve macaques. It was increased by 80% (1.86 day^{-1}) in the convalescent macaques vaccinated by αCD40.RBD or not.

The mechanistic model allows predicting the dynamics of unobserved compartments. Hence, a very early decrease of the target cells (all cells expressing ACE2) as well as of the viral inoculum which fully disappeared from day 2 onward were predicted (***Figure 2C***). In the three groups, the number of infected cells as well as infectious viral particles increased up to day 2 and then decreased. We show that this viral dynamic was blunted in the vaccinated animals leading to a predicted maximum number of infectious viral particles in the nasopharynx and the trachea below the detection threshold (***Figure 2C***). The number of target cells would be decreased by the infection in the naïve and the convalescent groups, whereas it would be preserved in vaccinated animals.

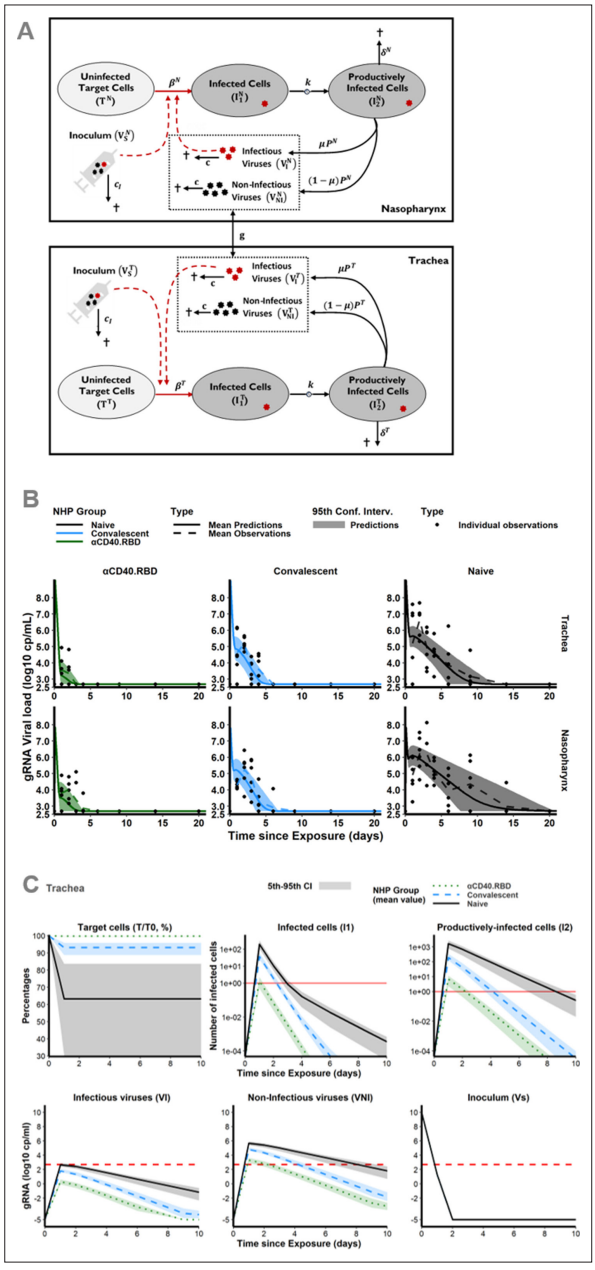


Figure 2. Mechanistic modelling. (**A**) Description of the model in the two compartments: the nasopharynx and the trachea. (**B**) Model fit to the $\log_{10}$ transformed observed genomic RNA (gRNA) viral loads in tracheal (top) and nasopharyngeal (bottom) compartments after the initial exposure to SARS-CoV-2 in naïve macaques (black, right) and after the second exposure in convalescent (blue, middle) and vaccinated (green, left) animals. Thick solid and

Figure 2 continued on next page

Figure 2 continued

dashed lines indicate mean viral load dynamics predicted and observed, respectively. Shaded areas indicate the 95% confidence intervals of the predictions. Dots represents observations. (**C**) Model predictions of unobserved quantities in the tracheal compartment for naïve (black, solid lines), convalescent (blue, dashed lines) and vaccinated (green, dotted lines) animals: target cells as percentage of the value at the challenge (top, left), infected cells (top, middle), productively infected cells (top, right), inoculum (bottom, right), infectious (bottom, left) and non-infectious virus (bottom, middle). Thick lines indicate mean values over time within each group. Shaded areas indicate the 95% confidence interval. Horizontal dashed red lines indicate the limit of quantification and horizontal solid red lines highlight the threshold of one infected cell.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Volumes of the trachea and nasopharynx, and weights measured at the time of exposure in four non-human primates (NHPs) in the study 1.

Source data 2. Weights of the 18 non-human primates (NHPs) in the study 1.

Source data 3. Genomic RNA (gRNA) viral load measured in the trachea and nasopharynx in the two additional non-human primates (NHPs) receiving inoculum via intra-gastric and intra-nasal routes.

Figure supplement 1. Modelling of the viral dynamics using mechanistic model.

Figure supplement 2. Modelling of the dynamics of viral replication.

The RBD-ACE2-binding inhibition is the main mechanistic CoP explaining the effect of the αCD40.RBD vaccine on new cell infection

In our study (***Marlin et al., 2021***), an extensive evaluation of the immunological response has been performed with quantification of spike-binding antibodies, antibodies inhibiting the attachment of RBD to ACE2, antibodies neutralizing infection, SARS-CoV-2-specific $CD4^+$ and $CD8^+$ T cells producing cytokines and serum cytokine levels (***Figure 3***, ***Figure 1—figure supplements 3–5***). Therefore, based on our mechanistic model, we investigated if any of these markers could serve as a mechanistic CoP. Such a CoP should be able to capture the effect of the natural immunity following

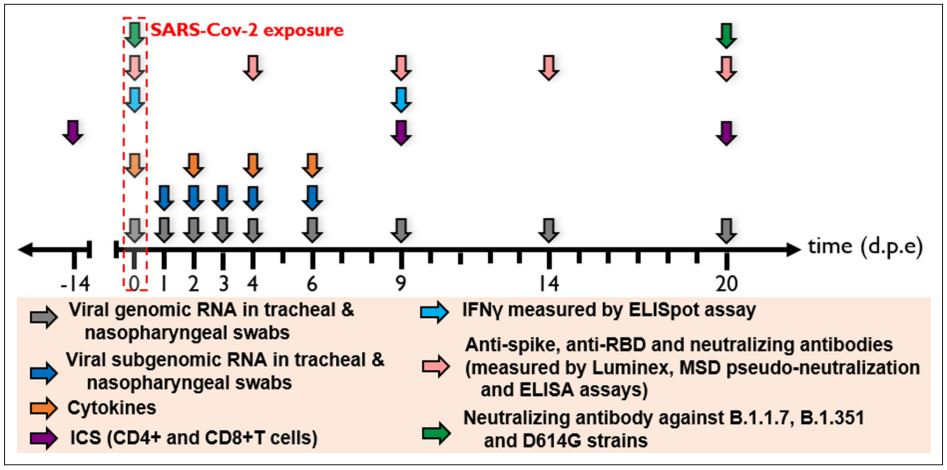


Figure 3. Harvest times and measurements. Nasopharyngeal and tracheal fluids were collected at 0, 1, 2, 3, 4, 6, 9, 14, and 20 days post-exposure (d.p.e) while blood was taken at 0, 2, 4, 6, 9, 14, and 20 d.p.e. Genomic and subgenomic viral loads were measured by RT-qPCR. Anti-spike IgG sera were titrated by multiplex bead assay, anti-RBD, and anti-nucleocapside (**N**) IgG were titrated using a commercially available multiplexed immunoassay developed by Mesoscale Discovery (MSD, Rockville, MD). The MSD pseudo-neutralization assay was used to measure antibodies neutralizing the binding of the spike protein and receptor-binding domain (RBD) to the ACE2 receptor. Neutralizing antibodies against B.1.1.7, B.1.351, and D614G strains were measured by S-Fuse neutralization assay and expressed as ED50 (effective dose 50%). T-cell responses were characterized as the frequency of PBMC expressing cytokines (IL-2, IL-17a, IFN-γ, TNF-α, IL-13, CD137, and CD154) after stimulation with S or N sequence overlapping peptide pools. IFN-γ ELISpot assay of PBMCs were performed on PBMC stimulated with RBD or N sequence overlapping peptide pools and expressed as spot-forming cell (SFC) per 1.0×10^6 PBMC.

infection, associated or not to the vaccine (group effect) estimated on both the rate of cell infection and the rate of the loss of infected cells. To this aim, we performed a systematic screening by adjusting the model for each marker and we compared these new models with the model without covariates and with the model adjusted for the groups. In particular, our approach allowed us to benefit from all the information provided by the overall dynamics of the immunological markers after the exposure by integrating them as time-varying covariates (see the Materials and methods section for a detailed description of the algorithm). We demonstrate that the RBD-ACE2-binding inhibition measure is sufficient to capture most of the effect of the groups on the infection of target cells (***Figure 4A and B***). The integration of this marker in the model explains the variability of the cell infection rate with greater certainty than the group of intervention, reducing the unexplained variability by 87% compared to 66% (***Supplementary file 1***). The marker actually takes into account the variation between animals within the same group. Hence, it suggests that the levels of anti-RBD antibodies induced by the vaccine that block attachment to ACE2 are highly efficient at reflecting the neutralization of new infections in vivo. Furthermore, when taking into account the information provided by the RBD-ACE2-binding inhibition assay, the effect of the group of intervention was no longer significant (***Supplementary file 1***). Finally, we looked at the estimated viral infectivity according to the binding inhibition assay in each animal. A positive dependence was found between the viral infectivity and the RBD-ACE2-binding inhibition measure, linking an increase of 10^3 AU of the marker, whether over time or between animals, with an increase of 1.8% (95CI% [1.2%; 2.3%]) of the viral infectivity (see ***Supplementary file 4***). Accordingly, the values at the time of exposure were not overlapping at all, distinguishing clearly the vaccinated and unvaccinated animals (see ***Figure 4C***).

In the next step, several markers (IgG-binding anti-RBD antibodies, $CD8^+$ T cells producing IFN-γ) appeared to be associated to the rate of loss of infected cells (***Figure 4—figure supplement 1A***). Both specific antibodies and specific $CD8^+$ T cells are mechanisms commonly considered important for killing infected cells. We retained the anti-RBD binding IgG Ab that were positively associated to the increase of the loss of infected cells. For unknown reason the IFN-γ response was high in unstimulated conditions in the naïve group. Thus, although this marker was associated with a decrease of the loss rate of infected cells, it appears essentially here as an indicator of the animal group. Further studies would be needed to fully confirm the place of IFN-γ response as a mechanistic marker.

A large part of the variation of the infection rate (71%) and loss rate of infected cells (60%) were captured by the two markers of CoP: the RBD-ACE2-binding inhibition and the anti-RBD-binding Ab concentration. Using the estimated parameters, the effective reproduction number could be calculated (R) which is representing the number of cells secondarily infected by virus from one infected cell (***Figure 4D***). When looking at this effective reproduction number according to the groups, the vaccinated animal presented from the first day of challenge an effective R below 1 meaning that no propagation of the infection started within the host. These results were consistent when taking the value of RBD-ACE2-binding inhibition at the time of the challenge without considering the evolution of the inhibition capacity over time (***Figure 4—figure supplement 1B***). This means that the dynamics of the viral replication is impacted very early during the infection process in immunized (i.e., both convalescent and vaccinated) animals and that vaccinated animals were protected from the beginning by the humoral response. Then, we looked at the threshold of the markers of interest leading to the control of the within-host infection (as defined by $R<1$) which was around 30,000 AU for the RBD-ACE2-binding inhibition assay. For the animals in the naïve and the convalescent groups, the observed values of binding inhibition measured by ECL RBD (the lower the better) and of IgG anti-RBD-binding antibodies (the higher the better) led to $R>1$, whereas in vaccinated animals, the value of ECL RBD led to $R<1$. Therefore, our modelling study shows that the inhibition of binding of RBD to ACE2 by antibodies is sufficient to control initial infection of the host (***Figure 4E***). According to the observed value of ECL RBD in vaccinated animals (e.g., 66 AU in ***Figure 4E***), a decrease of more than 2 $\log_{10}$ of the inhibition capacity (to reach 81,000 AU), due to variant of concern (VoC) or waning of immunity, would have been necessary to impair the control of the within-host infection. Moreover, a decrease of the neutralizing activity (i.e., increased ECL) could be compensated by an increase of cell death as measured by an increase of binding IgG anti-RBD as a surrogate. As an example, increasing IgG anti-RBD from 2.5 to 10 in the animal MF7 of the convalescent group would lead to a control of the infection.

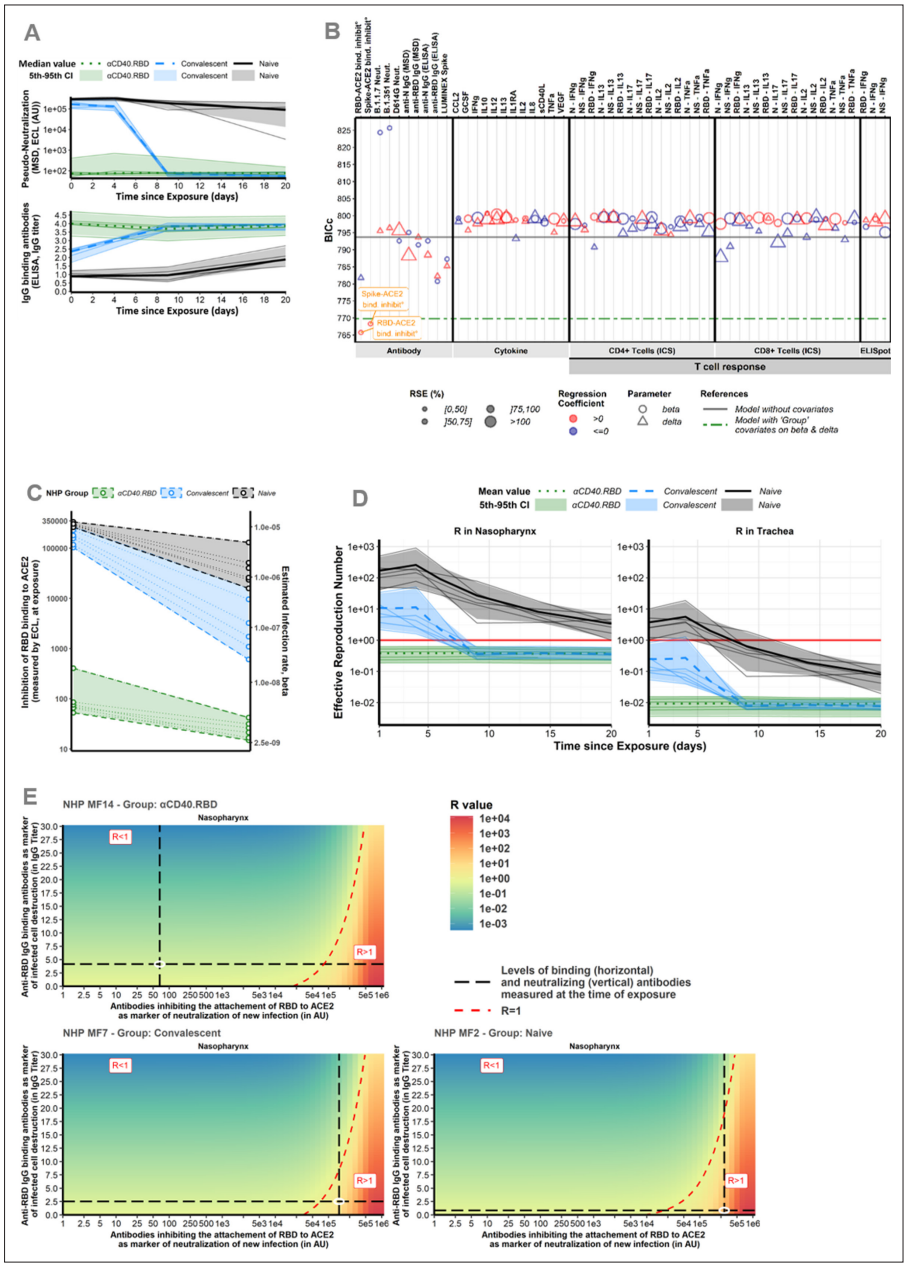


Figure 4. Immune markers. (**A**) *Dynamics of biomarker selected as mechanistic correlate of protection (mCoP).* Quantification of antibodies inhibiting RBD-ACE2 binding, measured by the Mesoscale Discovery (MSD) pseudo-neutralization assay (electro-chemiluminescence [ECL], in arbitrary unit [AU]) (top) and anti-RBD IgG titrated by ELISA assay (in IgG titer) (bottom). Thin lines represent individual values. Thick lines indicate medians of observations within naïve (black, solid line), convalescent (blue, dashed line), and αCD40.RBD-vaccinated convalescent (green, dotted line) animals. Shaded areas

Figure 4 continued on next page

Figure 4 continued

indicate 5th–95th confidence intervals of observations. (**B**) *Systematic screening of effect of the markers*. For every single marker, a model has been fitted to explore whether it explains the variation of the parameter of interest better or as well than the group indicator. Parameters of interest were β, the infection rate of ACE2+ target cells, and δ, the loss rate of infected cells. Models were compared according to the Bayesian information criterion (BIC), the lower being the better. The green line represents the reference model that includes the group effect (naïve/convalescent/vaccinated) without any adjustment for immunological marker (see ***Figure 3*** for more details about measurement of immunological markers). (**C**) *Thresholds of inhibition of RBD-ACE2 binding*. Estimated infection rate (in $(copies/mL)^{-1}$ day^{-1}) of target cells according to the quantification of antibodies inhibiting RBD-ACE2 (in ECL) at exposure. Thin dotted lines and circles represent individual values of infection rates (right axis) and neutralizing antibodies (left axis). Shaded areas delimit the pseudo-neutralization/viral infectivity relationships within each group. (**D**) *Reproduction number over time*. Model predictions of the reproduction number over time in the trachea (right) and nasopharynx (left). The reproduction number is representing the number of infected cells from one infected cell if target cells are unlimited. Below one, the effective reproduction number indicates that the infection is going to be cured. Horizontal solid red lines highlight the threshold of one. Same legend than (A). (**E**) *Conditions for controlling the infection*. Basic reproduction number (R_0) at the time of the challenge according to the levels of antibodies inhibiting RBD-ACE2 binding (the lower the better) and of anti-RBD IgG-binding antibodies (the higher the better) assuming they are mechanistic correlates of blocking new cell infection and promoting infected cell death, respectively. The red area with $R>1$ describes a situation where the infection is spreading. The green area with $R<1$ describes a situation where the infection is controlled. The dotted red line delimitates the two areas. Black long dashed lines represent the values of neutralizing and binding antibodies measured at exposure. Observed values for three different animals belonging to the naïve (bottom, right), convalescent (bottom, left), and vaccinated (top, left) groups are represented. For each animal, individual values of R_0 were estimated considering their individual values of the model parameters (β and δ).

The online version of this article includes the following source data and figure supplement(s) for figure 4:

Source data 1. Anti-N and anti-receptor-binding domain (RBD)-binding antibodies longitudinally measured post-immunization and quantified by ELISA in the study 1.

Source data 2. Anti-receptor-binding domain (RBD) and anti-spike neutralizing antibodies longitudinally measured post-exposition and quantified by Mesoscale Discovery (MSD) assay (in electro-chemiluminescence [ECL]) in the study 1.

Figure supplement 1. Immune markers selection and Basic reproduction number.

Figure supplement 2. Flowchart of the algorithm for automatic selection of covariate.

In conclusion, the αCD40.RBD vaccine-elicited humoral response leads to the blockade of new cell infection that is well captured by measure of the inhibition of attachment of the virus to ACE2 through the RBD of the spike protein. Hence, the inhibition of binding of RBD to ACE2 is a promising mechanistic CoP. Indeed, this CoP fulfills the three criteria of leading to the best fit (lower BIC), the best explanation of interindividual variability, and fully captured the effect of the group of intervention.

The model revealed the same CoP related to another protein-based vaccine but not with mRNA-1273 vaccine

We took the opportunity of another study testing a two-component spike nanoparticle protein-based vaccine performed in the same laboratory and using the same immune and virological assays (***Brouwer et al., 2021***), measured only at the time of exposure, for applying the proposed model and methodology. In this study, six animals were vaccinated and compared to four naïve animals (***Figure 5A and B***). The good fit of the data (***Figure 5C and D***) allows for estimating the effect of the vaccine that appeared here also to decrease the infectivity rate (by 99%) and increase the clearance of the infected cells by 79%. Looking at the best mechanistic CoP following the previously described strategy, we ended here again with the inhibition of RBD binding to ACE2 as measured by ECL RBD. In fact, this marker measured at baseline before challenge fulfilled the three criteria: (i) it led to the best model in front of a model adjusted for group effect, (ii) it rendered the group effect non-significant, and (iii) it explained around 71% of the infectivity rate variability, compared to 65% of variability explained by the groups. Interestingly, here again, the inhibition assay led to a clear separation of the estimated rate of infectivity between vaccinees and the placebo group (***Figure 5E***).

Finally, we applied our approach to a published NHP study performed to evaluate several doses of mRNA-1273 vaccine (***Corbett et al., 2020***). Using available data, we compared the viral dynamics in the 100 µg, 10 µg, and placebo groups, enrolling a total of 12 rhesus macaques in a 1:1:1 ratio. Similar to the previous study, only immune markers measured at the time of exposure were available in this study, in addition to viral dynamics. We started from the same model as defined previously. We estimated a reduction of the infection rate by 97% but we did not find any additional effect. Looking at potential mechanistic CoP, we retained neutralization as measured on live cells with Luciferase marker. Although this marker led to the best fit and replaced the group effect (which was non-significant after

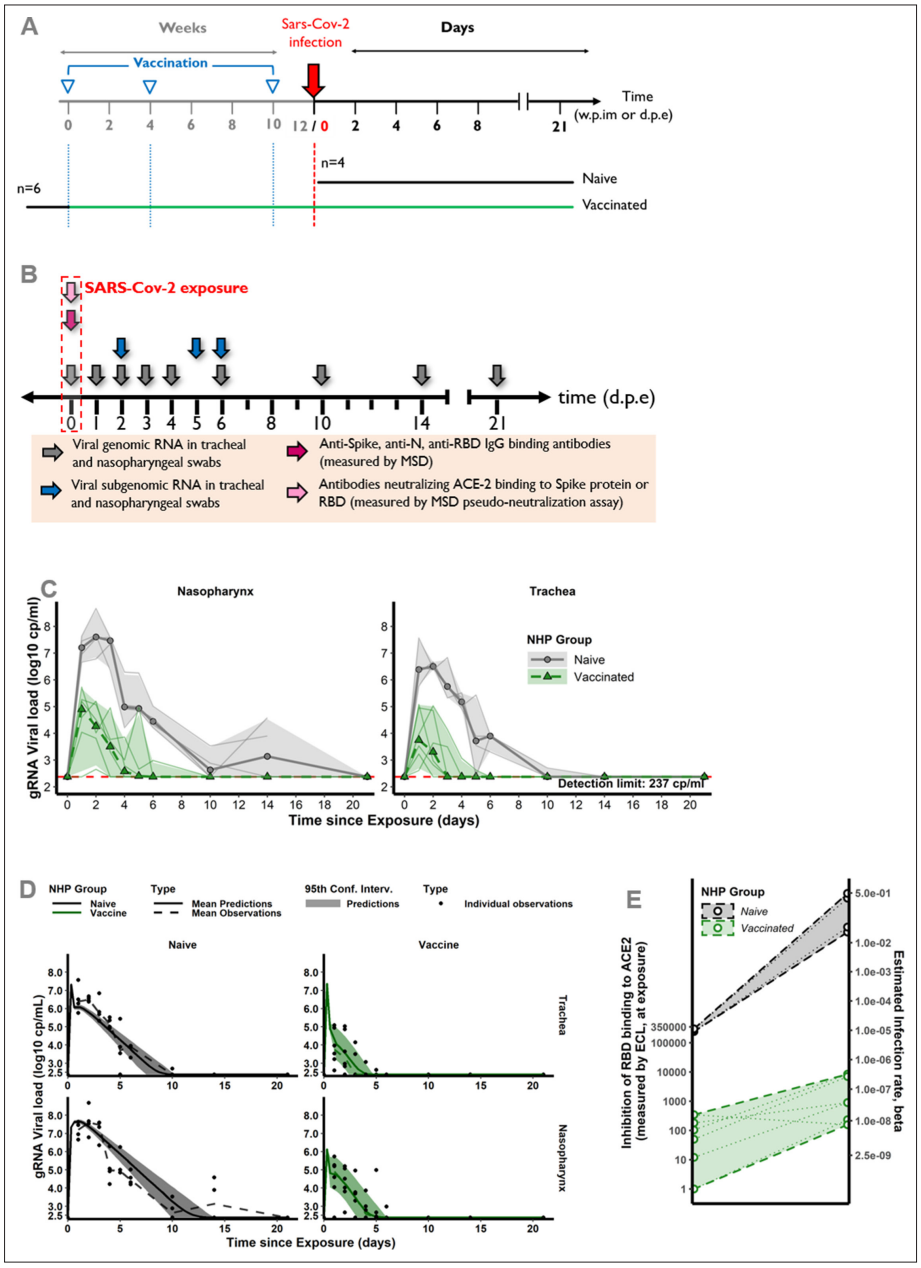


Figure 5. Study design and modeling results for the second study testing two-component spike nanoparticle vaccine.
(**A**) Study design. Cynomolgus macaques were randomly assigned in two experimental groups. Twelve, eight, and two weeks post-infection with SARS-CoV-2 virus, six of them were successively immunized with 50 µg of SARS-CoV-2 S-I53-50NP vaccine. The four other animals received no vaccination. Two weeks after the final immunization, all monkeys were exposed to a total dose of 10^6 pfu of SARS-CoV-2 virus via intra-nasal and intra-tracheal routes.

Figure 5 continued on next page

Figure 5 continued

(**B**) Harvest times and measurements. Nasopharyngeal and tracheal fluids were collected at 0, 1, 2, 3, 4, 5, 6, 10, 14, and 21 days post-exposure (d.p.e.) while blood was taken at 0, 2, 4, 6, 10, 14, and 21 d.p.e. Genomic and subgenomic viral loads were measured by RT-qPCR. Anti-spike, anti-RBD, and anti-nucleocapside (**N**) IgG were titrated using a multiplexed immunoassay developed by Mesoscale Discovery (MSD, Rockville, MD) and expressed in AU mL^{-1}. The MSD pseudo-neutralization assay was used to quantify antibodies neutralizing the binding of the spike protein and RBD to the ACE2 receptor and results were expressed in electro-chemiluminescence (ECL). (**C**) Genomic viral load dynamics in nasopharyngeal and tracheal swabs after the exposure to SARS-Cov-2 in naïve (black, solid line) and vaccinated (green, dashed line) animals. Thin lines represent individual values. Thick lines indicate medians within each group. (**D**) Model fit to the log_{10}-transformed observed genomic RNA (gRNA) viral load in nasopharynx and trachea after the exposure to SARS-CoV-2 in naïve and vaccinated macaques. Solid thin lines indicate individual dynamics predicted by the model adjusted for groups. Thick dashed lines indicate mean viral load over time. (**E**) Thresholds of inhibition of RBD-ACE2 binding. Estimated infection rate of target cells $((copies/mL)^{-1}\ day^{-1})$ according to the quantification of antibodies inhibiting RBD-ACE2 binding (ECL) at exposure for naïve (black) and vaccinated (green) animals. Thin dotted lines and circles represent individual infection rates (right axis) and neutralizing antibodies (left axis). Thick dashed lines and dashed areas delimit the pseudo-neutralization/viral infectivity relationships within each group. (**C,D**) Horizontal red dashed lines represent the limit of quantification and shaded areas the 95% confidence intervals.The second study testing two-component spike nanoparticle vaccine.

The online version of this article includes the following source data for figure 5:

Source data 1. Anti-spike, anti-receptor-binding domain (RBD), and anti-N-binding antibodies quantified by Mesoscale Discovery (MSD) assay (AU mL^{-1}), and quantification of the spike/ACE2-binding inhibition by MSD assay (in 1/ECL), at the time of exposure in the study 2.

Source data 2. Genomic RNA (gRNA) and subgenomic RNA (sgRNA) viral loads longitudinally measured in the trachea and nasopharynx in the study 2.

adjustment for the marker), it explained only 15% of the variability of estimated viral infectivity, while 19% were explained by the groups.

In conclusion, we demonstrated, based upon challenge studies in NHP vaccinated with two different protein-based vaccine platforms, that both vaccines lead to the blockade of new cell infection. Neutralizing antibodies likely represent a consistent mechanistic correlate of protection (mCoP). This could change across vaccine platforms especially because mechanisms of action are different.

Discussion

We explored the mechanistic effects of three SARS-CoV-2 vaccines and assessed the quality of markers as mCoP. This model showed that neutralizing and binding antibodies elicited by a non-adjuvanted protein-based vaccine targeting the RBD of spike to the CD40 receptor of antigen presenting cells are reliable mCoP. Interestingly, we found the simpler and easier to standardize and implement binding inhibition assay may be more relevant to use as a CoP than cell-culture neutralization assays. This result has been replicated in another study testing a nanoparticle spike vaccine. The model was able to capture the effect of the vaccines on the reduction of the rate of infection of target cells and identified additional effects of vaccines beyond neutralizing antibodies. This latter consisted of increasing the loss rate of infected cells which was better reflected by the IgG-binding antibodies and $CD8^+$ T-cell responses in the case of the CD40-targeting vaccine. One limitation of our study is that the prediction potential of our model relies on the range of the immune markers measured. However, our approach would allow a full exploitation of the data generated as in systems serology where non-neutralizing Ab functions, such as ADCC, ADCP, ADCD, and Ab-dependent respiratory burst (ADRB) are explored (***Chung et al., 2015***). The role of ADCC in natural infection has been previously shown (***Dufloo et al., 2021***), ADCD in DNA vaccine recipients (***Yu et al., 2020***) and with Ad26 vaccine (***Alter et al., 2021***). Here, we extended significantly these data by modelling the viral dynamic, showing that two other protein-based vaccines exert an additional effect on infected cell death which relied on the level of IgG anti-RBD-binding antibodies especially for the CD40.RBD-targeting vaccine. Measurements of other non-neutralizing Ab functions would probably also capture this additional effect.

The next question after determining which marker is a valid mCoP is to define the concentration that leads to protection, looking for a threshold effect that will help to define an objective (***Khoury et al., 2021***; ***Jin et al., 2021***). In the context of SARS-CoV-2 virus, several emerged variants are leading to a significant reduction of viral neutralization as measured by various approaches. However, a 20-fold reduction of viral neutralization might not translate in 20-fold reduction of vaccine efficacy (***Emary et al., 2021***). First, there are many steps between viral neutralization and the reduction of viral infectivity or the improvement of clinical symptoms. Second, the consequences of a reduction of viral neutralization could be alleviated by other immunological mechanisms not compromised by

the variant. In the context of natural immunity, when the level of neutralizing antibodies was below a protective threshold, the cellular immune response appeared to be critical (**McMahan et al., 2021**; **Chandrashekar et al., 2020**). We showed with our model that an improvement of infected cell destruction could help to control the within-host infection and is quantitatively feasible.

The control of viral replication is the key for reducing infectivity (**Leung et al., 2020**; **Marks et al., 2021**) as well as disease severity (**Néant, 2021**; **Gutmann et al., 2021**). According to our non-linear model linking the neutralization to the viral replication, a decrease of 4- to 20-fold in neutralization as described for the variants of concern (**Planas et al., 2021**; **Zhou et al., 2021**) is not enough, especially in the context of the response to CD40.RBD-targeting vaccine, to compromise the control of viral replication. The results showing a conserved effectiveness of mRNA vaccines in humans infected by the alpha or beta variants (**Charmet et al., 2021**), although a decrease of neutralization has been reported (**Planas et al., 2021**), are consistent with this hypothesis. However, this is highly dependent upon the mode of action of currently used vaccines and upon the VoC that may much more compromise the neutralization but being also intrinsically less pathogenic such as Omicron (**Nyberg et al., 2022**).

The analysis performed extended significantly the observation of associations between markers as previously reported for SARS-CoV-2 vaccine (**Yu et al., 2020**) and other vaccines (**Kester et al., 2009**) because it allows a more causal interpretation of the effect of immune markers. However, our modelling approach requires the in vivo identification of the biological parameters under specific experimentations. On the other hand, the estimation of parameters included in our model also provided information on some aspect of the virus pathophysiology. Notably, we found an increased capacity of virion production in nasopharynx compared to the trachea which could be explained by the difference in target cells according to the compartment (**Travaglini et al., 2020**). This result needs to be confirmed as it may also be the consequence of a different local immune response (**Pizzorno et al., 2020**). The choice of the structural model defining the host-pathogen interaction is a fundamental step in the presented approach. Here, it was well guided by the biological knowledge, the existing models for viral dynamics (**Goyal et al., 2019**; **Gonçalves et al., 2021**; **Smith et al., 2018**), and the statistical inference allowing the selection of the model that best fit the data. As the number of observations was relatively small in regard to the number of model parameters, we investigated overfitting issues. This was done using a bootstrap approach to evaluate the stability of confidence intervals of the estimated parameters. Results are provided in Appendix 2 'BICc as selection criteria and multiple testing adjustment'. Many modelling choices for the statistical model were made in this approach and more theoretical work evaluating the robustness of the results in their regards may be relevant for future works. In particular, we could relax the constraint of linear interpolation of marker dynamics by using simple regression models, allowing in the same time the integration of error model to account for measurement error for time-varying covariates (**Dafni and Tsiatis, 1998**; **Carroll et al., 2006**; **Wu, 2009**). Moreover, by construction, we assumed similar interindividual variability and effects of covariates within the two URT compartments as well as similar values for the viral infectivity and the loss rate of infected cells. Viral load dynamics measured in lungs being different from those in the URT (**Lui et al., 2020**; **Goyal et al., 2020**), the relaxation of this hypothesis of homogeneous physiological behavior in the URT may be pertinent to extend the model to the LRT. Finally, it should be underlined that the dynamics of the immune response has not been modelled as suggested for instance for B-cell response (**Balelli et al., 2020**). This clearly constitutes the next step after the selection of the markers of interest as done in the present work.

In conclusion, the modelling of the response to two new promising SARS-CoV-2 vaccines in NHP revealed a combination of effects with a blockade of new cell infections and the destruction of infected cells. For these two vaccines, the antibody inhibiting the attachment of RBD to ACE2 appeared to be a very good surrogate of the vaccine effect on the rate of infection of new cells and therefore could be used as a mechanistic CoP. This modelling framework contributes to the improvement of the understanding of the immunological concepts by adding a quantitative evaluation of the contributions of different mechanisms of control of viral infection. In terms of acceleration of vaccine development, our results may help to develop vaccines for 'hard-to-target pathogens', or to predict their efficacy in aging and particular populations (**Pollard and Bijker, 2021**). It should also help in choosing vaccine dose, for instance at early development (**Rhodes et al., 2018**) as well as deciding if and when boosting vaccination is needed in the face of waning protective antibody levels (**Gaebler et al., 2021**;

Vanshylla et al., 2021), at least in NHP studies although the framework could be extended to human studies using mixed approaches of within and between hosts modelling (*Goyal et al., 2022*) providing that enough information is collected.

Materials and methods

Experimental model and subjects details

Cynomolgus macaques (*Macaca fascicularis*), aged 37–66 months (18 females and 13 males) and originating from Mauritian AAALAC certified breeding centers were used in this study. All animals were housed in IDMIT facilities (CEA, Fontenay-aux-roses), under BSL2 and BSL-3 containment when necessary (Animal facility authorization #D92-032-02, Préfecture des Hauts de Seine, France) and in compliance with European Directive 2010/63/EU, the French regulations and the Standards for Human Care and Use of Laboratory Animals, of the Office for Laboratory Animal Welfare (OLAW, assurance number #A5826-01, US). The protocols were approved by the institutional ethical committee 'Comité d'Ethique en Expérimentation Animale du Commissariat à l'Energie Atomique et aux Energies Alternatives' (CEtEA #44) under statement number A20-011. The study was authorized by the 'Research, Innovation and Education Ministry' under registration number APAFIS#24434-2020030216532863v1.

Evaluation of anti-spike, anti-RBD, and neutralizing IgG antibodies

Anti-spike IgG were titrated by multiplex bead assay. Briefly, Luminex beads were coupled to the spike protein as previously described (*Fenwick et al., 2021*) and added to a Bio-Plex plate (Bio-Rad). Beads were washed with PBS 0.05% tween using a magnetic plate washer (MAG2x program) and incubated for 1 hr with serial diluted individual serum. Beads were then washed and anti-NHP IgG-PE secondary antibody (Southern Biotech, clone SB108a) was added at a 1:500 dilution for 45 min at room temperature (RT). After washing, beads were resuspended in a reading buffer 5 min under agitation (800 rpm) on the plate shaker then read directly on a Luminex Bioplex 200 plate reader (Bio-Rad). Average MFI from the baseline samples were used as reference value for the negative control. Amount of anti-spike IgG was reported as the MFI signal divided by the mean signal for the negative controls.

Anti-RBD and anti-nucleocapside (N) IgG were titrated using a commercially available multiplexed immunoassay developed by Mesoscale Discovery (MSD, Rockville, MD) as previously described (*Johnson et al., 2020*). Briefly, antigens were spotted at 200–400 µg mL^{-1} in a proprietary buffer, washed, dried, and packaged for further use (MSD Coronavirus Plate 2). Then, plates were blocked with MSD Blocker A following which reference standard, controls, and samples diluted 1:500 and 1:5000 in diluent buffer were added. After incubation, detection antibody was added (MSD SULFO-TAGTM Anti-Human IgG Antibody) and then MSD GOLDTM Read Buffer B was added and plates read using a MESO QuickPlex SQ 120 MM Reader. Results were expressed as arbitrary unit (AU) mL^{-1}.

Anti-RBD and anti-N IgG were titrated by ELISA. The nucleocapsid and the spike RBD (Genbank # NC_045512.2) were cloned and produced in *Escherichia coli* and CHO cells, respectively, as previously described (*Flamar et al., 2012*). Antigens were purified on C-tag column (Thermo Fisher) and quality-controlled by SDS-PAGE and for their level of endotoxin. Antigens were coated in a 96-well plates Nunc-immuno Maxisorp (Thermo Fisher) at 1 µg mL^{-1} in carbonate buffer at 4°C overnight. Plates were washed in TBS Tween 0.05% (Thermo Fisher) and blocked with PBS 3% BSA for 2 hr at RT. Samples were then added, in duplicate, in serial dilution for 1 hr at RT. Non-infected NHP sera were used as negative controls. After washing, anti-NHP IgG coupled with HRP (Thermo Fisher) was added at 1:20,000 for 45 min at RT. After washing, TMB substrate (Thermo Fisher) was added for 15 min at RT and the reaction was stopped with 1 M sulfuric acid. Absorbance of each well was measured at 450 nm (reference 570 nm) using a Tristar2 reader (Berthold Technologies). The EC50 value of each sample was determined using GraphPad Prism 8 and antibody titer was calculated as log (1/EC50).

The MSD pseudo-neutralization assay was used to measure antibodies neutralizing the binding of the spike protein to the ACE2 receptor. Plates were blocked and washed as above, assay calibrator (COVID- 19 neutralizing antibody; monoclonal antibody against S protein; 200 µg mL^{-1}), control sera, and test sera samples diluted 1:10 and 1:100 in assay diluent were added to the plates. Following incubation of the plates, an 0.25 µg mL^{-1} solution of MSD SULFO-TAGTM-conjugated ACE2 was added after which plates were read as above. Electro-chemiluminescence (ECL) signal was recorded.

Viral dynamics modelling

The mechanistic approach we developed to characterize the impact of the immune response on the viral gRNA and sgRNA dynamics relies on a mechanistic model divided in three layers: first, we used a mathematical model based on ODEs to describe the dynamics in the two compartments, the nasopharynx and the trachea. Then, we used a statistical model to take into account both the interindividual variability and the effects of covariates on parameters. Finally, we considered an observation model to describe the observed $\log_{10}$ viral loads in the two compartments.

For the mathematical model, we started from previously published models (***Gonçalves et al., 2020***; ***Kim et al., 2021***; ***Baccam et al., 2006***) where the nasopharynx and trachea were respectively described by a target cell limited model, with an eclipse phase, as model of acute viral infection assuming target-cell limitation (***Baccam et al., 2006***). We completed the model by adding a compartment for the inoculum that distinguishes the injected virus (V_s) from the virus produced de novo (V_i and V_{ni}). To our knowledge, this distinction has not been proposed in any previous work. Two main reasons led us to make this choice. First, it allowed us to study the dynamics of the inoculum, in particular during the early phase of viral RNA load dynamics. Second, as described in more detail below, it gave us the opportunity to use all the information provided by the preclinical studies, such as the known number of inoculated virions, to define the initial conditions of the ODE model rather than estimating or randomly fixing them for V_i and V_{ni}, as is usually done. Consequently, for each of the two compartments, the model included uninfected target cells (T) that can be infected (I_1) either by infectious viruses (V_i) or inoculum (V_s) at an infection rate β. After an eclipse phase, infected cells become productively infected cells (I_2) and can produce virions at rate P and be lost at a per capita rate δ. The virions generated can be infectious (V_i) with proportion μ while the $(1-\mu)$ remaining proportion of virions is non-infectious (V_{ni}). Mathematically, a single compartment (V) for de novo produced virions could be considered in the model, with μV and $(1-\mu)V$ representing the respective contributions of infectious and non-infectious viruses to the biological mechanisms. However, to have a better visual understanding of the distinction between the two types of viruses, we wrote the model with distinct compartments, V_i and V_{ni}.

Finally, virions produced de novo and those from the inoculum are cleared at a rate c and c_i, respectively. Distinct clearances were considered to account for the effects of experimental conditions on viral dynamics. In particular, it is hypothesized that, animals being locally infected with large numbers of virions, a large proportion of it is assumed to be rapidly eliminated by swallowing and natural downstream influx, in contrast to the de novo-produced virions. However, it is important to keep in mind that this distinction was possible because of the controlled experimental conditions performed in animals, (i.e., exact timing and amount of inoculated virus known, and frequent monitoring during the early phase of the viral dynamics). Because of identifiability issues, similar clearances for infectious and non-infectious viruses were used. Accordingly, the model can be written as the following set of differential equations, where the superscript X denotes the compartment of interest (N, nasopharynx or T, trachea):

$$\begin{cases} \frac{dT^X}{dt} = -\beta^X V_i^X T^X - \mu\beta^X V_s^X T^X \\ \frac{dI_1^X}{dt} = \beta^X V_i^X T^X + \mu\beta^X V_s^X T^X - kI_1^X \\ \frac{dI_2^X}{dt} = kI_1^X - \delta^X I_2^X \\ \frac{dV_i^X}{dt} = \mu P^X I_2^X - cV_i^X - \beta^X V_i^X T^X \\ \frac{dV_{ni}^X}{dt} = (1-\mu) P^X I_2^X - cV_{ni}^X \\ \frac{dV_s^X}{dt} = -c_i V_s^X - \mu\beta^X V_s^X T^X \\ \\ T^X(t=0) = T_0^X ; I_1^X(t=0) = 0 ; I_2^X(t=0) = 0 \\ V_i^X(t=0) = 0 ; V_{ni}^X(t=0) = 0 ; V_s^X(t=0) = V_{S,0}^X \end{cases} \quad (1)$$

where $T^X(t=0)$, $I_1^X(t=0)$, $I_2^X(t=0)$, $V_i^X(t=0)$, $V_{ni}^X(t=0)$, and $V_s^X(t=0)$ are the initial conditions at the time of exposure. The initial concentration of target cells, that are the epithelial cells expressing the ACE2 receptor, is expressed as $T_0^X = \frac{T_0^{X,nbc}}{W^X}$, where $T_0^{X,nbc}$ is the initial number of cells and W^X is

the volume of distribution of the compartment of interest (see the subsection 'Consideration of the volume of distribution'). Each animal was exposed to 1 × 10^6 pfu of SARS-CoV-2 representing a total of 2.19 × 10^{10} virions. Over the total inoculum injected (5 mL), 10% (0.5 mL) and 90% (4.5 mL) of virions were respectively injected by the intra-nasal route and the intra-tracheal route leading to the following initial concentrations of the inoculum within each compartment: $V_{S,0}^N = \frac{0.10 \times Inoc_0}{W^N}$ and $V_{S,0}^T = \frac{0.90 \times \mathrm{Inoc}_0}{W^T}$, with $Inoc_0$ the number of virions injected via the inoculum.

Using the gRNA and sgRNA viral loads, we estimated the viral infectivity, the viral production rate, and the loss rate of infected cells within each of the two compartments of the URT (***Supplementary file 2***). To account for interindividual variability and covariates, each of those three parameters was described by a mixed-effect model and jointly estimated between the two compartments as follows:

$$\begin{cases} \log_{10}\left(\beta_i^N\right) = \beta_0 + \phi_{conv}^{\beta} \times I_{group=conv} + \phi_{CD40}^{\beta} \times I_{group=CD40} + u_i^{\beta} \\ \beta_i^T = \beta_i^N \times \exp\left(f_{\beta}^T\right) \\ \log\left(\delta_i^N\right) = \log\left(\delta_0\right) + \phi_{conv}^{\delta} \times I_{group=conv} + \phi_{CD40}^{\delta} \times I_{group=CD40} + u_i^{\delta} \\ \delta_i^T = \delta_i^N \times \exp\left(f_{\delta}^T\right) \\ \log\left(P_i^N\right) = \log\left(P_0\right) + \phi_{conv}^{P} \times I_{group=conv} + \phi_{CD40}^{P} \times I_{group=CD40} + u_i^{P} \\ P_i^T = P_i^N \times \exp\left(f_P^T\right) \end{cases} \quad (2)$$

where β_0, $\log\left(\delta_0\right)$, and $\log\left(P_0\right)$ are the fixed effects, $\left\{\phi_{conv}^{\theta} \mid \theta \in \{\beta, \delta, P\}\right\}$ and $\left\{\phi_{CD40}^{\theta} \mid \theta \in \{\beta, \delta, P\}\right\}$ are respectively the regression coefficients related to the effects of the group of convalescent and αCD40.RBD-vaccinated animals for the parameters β, δ, and P, and u_i^{θ} is the individual random effect for the parameter θ, which is assumed to be normally distributed with variance ω_{θ}^2 . A log-transformation was adopted for the parameters δ and P to ensure their positivity while a $\log_{10}$-transformation was chosen for viral infectivity to also improve the convergence of the estimation. Because of the scale difference between the parameter β and the other parameters (see ***Supplementary file 2***), the mere use of the log-transformation for this parameter led to convergence issues. The use of a $\log_{10}$-transformation allowed to overcome this problem. Moreover, as shown in ***Equation 2***, a joint estimation of the parameters β, δ, and P between the two compartments of the URT was considered. In this regard, a homogeneous interindividual variability within the URT was assumed as well as a similar contribution of the covariates to the value of the parameters. Parameters in the trachea were then either equal or proportional to those in the nasopharynx. This modelling choice, resulting in a smaller number of parameters to be estimated, was made mainly to address identifiability issues and to increase the power of the estimation. All other parameters included in the target-cell limited models were assumed to be fixed (see the subsection 'Parameter estimation' for more details).

In practice, after the selection of the optimal statistical model (see Appendix 1 'Model building'), random effects were added only to the parameters β and δ (i.e., $\omega_{\beta} \neq 0$, $\omega_{\delta} \neq 0$, and $\omega_P = 0$), and the estimation of multiple models identified the viral production rate P as the only parameter taking different values between the trachea and nasopharynx. (i.e., $\beta^N = \beta^T$ with $f_{\beta}^T = 0$, $\delta^N = \delta^T$ with $f_{\delta}^T = 0$, while $P^N \neq P^T$). Finally, the adjustment of the model for the categorical covariates of groups of treatment, natural infection, and/or vaccination identified β and δ as the parameters with a statistically significant effect of these covariates (i.e., $\phi_{conv}^P = 0$ and $\phi_{CD40}^P = 0$).

For the observation model, we jointly described genomic and subgenomic viral loads in the two compartments of the URT. We defined genomic viral load, which characterizes the total viral load observed in a compartment (nasopharynx or trachea), as the sum of inoculated virions (V_s), infectious (V_i), and non-infectious virions (V_{ni}). The sgRNA was described as proportional to the infected cells ($I_1 + I_2$). This choice was driven by two main reasons. First, sgRNA is only transcribed in infected cells (***Sawicki et al., 2007***). Second, as described by ***Miao et al., 2011***, to overcome identifiability issues between the parameters β and P typically observed in target-cell limited models. The comparison of the two observation models describing sgRNA as either proportional to virions produced de novo ($V_i + V_{ni}$) or proportional to infected cells ($I_1 + I_2$) confirmed this conclusion. In addition to a better BICc value (–25 points) compared with the first model, the second one allowed the estimation of both β and P by counteracting identifiability problems faced with the first model (results not shown). Accordingly, the $\log_{10}$-transformed gRNA and sgRNA of the ith animal at the jth time point in compartment X

(nasopharynx or trachea), denoted $gRNA_{ij}^{X}$ and $sgRNA_{ij}^{X}$, respectively, were described by the following equations:

$$\begin{cases} gRNA_{ij}^{X} = \log_{10}\left[\left(V_i^X + V_{ni}^X + V_s^X\right)\left(\Theta_i^X, t_{ij}\right)\right] + \varepsilon_{ij,g}^X \quad \varepsilon_{ij,g}^X \sim N\left(0, \sigma_{gX}^2\right) \\ sgRNA_{ij}^{X} = \alpha_{sgRNA} \times \log_{10}\left[\left(I_1^X + I_2^X\right)\left(\Theta_i^X, t_{ij}\right)\right] + \varepsilon_{ij,sg}^X \quad \varepsilon_{ij,sg}^X \sim N\left(0, \sigma_{sgX}^2\right) \end{cases} \tag{3}$$

where Θ_i^X is the set of parameters of the subject *i* for the compartment *X* and ε are the additive normally distributed measurement errors.

Consideration of the volume of distribution

To define the concentration of inoculum within each compartment after injection, nasopharyngeal and tracheal volumes of distribution, labelled W^N and W^T, respectively, were needed. Given the estimated volumes of the trachea and the nasal cavities in four monkeys similar to our 18 macaques (***Figure 2—figure supplement 2A–C***) and the well-documented relationship between the volume of respiratory tract and animal weights (***Asgharian et al., 2012***), the volume of distribution of each compartment was defined as a step function of NHP weights:

$$\begin{aligned} W_i^N &= \begin{cases} 4 & \text{if weight}_i \leq 4.5 \\ 5.5 & \text{otherwise} \end{cases} \\ W_i^N &= \begin{cases} 2 & \text{if weight}_i \leq 4.5 \\ 3 & \text{otherwise} \end{cases} \end{aligned} \tag{4}$$

where weight_i is the weight of the monkey *i* in kg. Using ***Equation 4*** and weights of our 18 NHPs (mean = 4.08; [Q1; Q3] = [3.26; 4.77]), we estimated WT = 2 and WN = 4 mL for a third of them (n=12) (***Figure 2—figure supplement 2D***), leading to the initial concentration of target cells T_0^X (see 'Viral dynamics modelling' for equation) fixed at 3.13×10^4 cells mL^{-1} and 1.13×10^4 cells mL^{-1} in nasopharynx and trachea, respectively. Similarly, their initial concentrations of challenge inoculum $V_{S,0}^X$ were fixed at 5.48×10^8 copies,mL^{-1} and 9.86×10^9 copies,mL^{-1} in nasopharynx and trachea respectively. For the last third of NHPs (*n*=6), WT = 3 and WN = 5.5 mL leading to T_0^X fixed at 2.27×10^4 cells mL^{-1} in nasopharynx and 7.50×10^3 cells mL^{-1} in trachea while $V_{S,0}^X$ was fixed at 3.98×10^8 copies mL^{-1} in nasopharynx and 6.57×10^9 copies mL^{-1} in trachea. Through this modelling, we assumed a homogenous distribution of injected virions and target cells within nasopharyngeal and tracheal compartments. In addition, the natural downward flow of inoculum toward lungs, at the moment of injection, was indirectly taken into account by the parameter of inoculum clearance, c_i.

Parameter estimation

Among all parameters involved in the three layers of the mechanistic model, some of them have been fixed based on experimental settings and/or literature. That is the case of the proportion of infectious virus (μ) that has been fixed at 1/1000 according to previous work (***Gonçalves et al., 2021***) and additional work (results not shown) evaluating the stability of the model estimation according to the value of this parameter. The initial number of target cells, that are the epithelial cells expressing the ACE2 receptor, $T_0^{X,nbc}$ was fixed at 1.25×10^5 cells in the nasopharynx and 2.25×10^4 cells in trachea (***Gonçalves et al., 2021***; ***Supplementary file 2***). The duration of the eclipse phase ($1/k$), the clearance of the inoculum (c_i) and the clearance of the virus produced de novo (c) were estimated by profile likelihood. The profile likelihood consists in defining a grid of values for the parameters to be evaluated and sequentially fixing these parameters to one of these combinations of values. The model and all the parameters that are not fixed are then estimated by maximizing the log-likelihood. In this process, all parameters that are assumed to be fixed in the model (i.e., μ and the initial conditions) are held fixed. Finally, the optimal set of parameters is chosen as the one optimizing the log-likelihood. Although the available data did not allow the direct estimation of these three parameters, the use profile likelihood enabled the exploration of various potential values for k, c, and c_i . In a first step, we explored the 18 models resulting from the combination of three values of $k \in \{1, 3, 6\}$ day^{-1} and six values for $c \in \{1, 5, 10, 15, 20, 30\}$ day^{-1}, assuming that the two parameters of virus clearance were

equal, as first approximation. As shown in ***Supplementary file 3a***, an eclipse phase of 8 hr (k=3) and virus clearance higher than 15 virions per day led to lowest values of –2log-likelihood (–2LL, the lower the better). In a second step, we fixed the parameter k at 3 day^{-1} and estimated the 70 models resulting from the combination of 10 values for $c\in\{1, 2, 3, 4, 5, 10, 15, 20, 25, 30\}$ day^{-1} and 7 values for$\in\{1, 5, 10, 15, 20, 25, 30\}$ day^{-1} (***Supplementary file 3b***). The distinction of the two parameters of free virus clearance enabled to find much lower half-life of inoculum (~50 min) than half-life of virus produced de novo (~5.55 hr), with c=3 day^{-1} compared to c_i = 20 day^{-1}.

Once all these parameters have been fixed, the estimation problem was restricted to the determination of the viral infectivity β, the viral production rate P, the loss rate of infected cells δ for each compartment, the parameter α_{vlsg} in the observation model, regression coefficients for groups of intervention $(\phi_{conv}, \phi_{CD40})$, and standard deviations for both random effects (ω) and error model (σ). The estimation was performed by maximum likelihood estimation using a stochastic approximation EM algorithm implemented in the software Monolix (http://www.lixoft.com). The Fisher information matrix was calculated by stochastic approximation, providing for each estimated parameter its variance, from which we were able to derive its 95% confidence interval. Selection of the compartment effect on parameters (β, δ, P) as well as random effects and covariates on the statistical model (***Equation 2***) was performed by the estimation of several models that were successively compared according to the corrected Bayesian information criterion (BICc) (to be minimized). After the removal of random effect on the viral production ($\omega_P = 0$) allowing the reduction of the variance on the two other random effects, all combinations of compartment effects were evaluated, leading to the final selection of a single effect on $P\left(f_\beta^T = f_\delta^T = 0\right)$. Then, the effect of group intervention was independently added on model parameters among β, δ, P, and c. Once the group effect on the viral infectivity identified as the best one, the addition of a second effect on the remaining parameters was tested, resulting in the selection of the loss rate of infected cells. Finally, the irrelevance of the addition of a third effect was verified.

The possibility of migration of free plasma virus between the nasopharynx and the trachea was tested. However, as widely described in the literature, the transport of viral particles within the respiratory tract is negligible in the viral dynamics and is difficult to estimate. The reader can refer to Appendix 1 'Model building' for an additional modelling work conducted to estimate this exchange and provided the same conclusion. Accordingly, the two compartments of the URT were assumed are distinct in our model.

Algorithm for automatic selection of biomarkers as CoP

After identifying the effect of the group of intervention on both the viral infectivity (β) and the loss rate of infected cells (δ), we aimed at determining whether some immunological markers quantified in the study could capture this effect. Nowadays, many methods for selecting constant covariates already exist (***Chowdhury and Turin, 2020***) and are implemented in software like Monolix. However, these latter do not allow time-varying covariates. In this section, we present the algorithm we implemented to select time-varying covariates. We proposed a classical stepwise data-driven automatic covariate modelling method (***Figure 4—figure supplement 2***). However, initially implemented to select covariates from more than 50 biomarkers, computational time restricted us to consider only a forward selection procedure. Nevertheless, the method can be easily extended to classical stepwise selection in which both forward selection and backward elimination are performed sequentially. Although the method was developed for time-varying covariates, it can also be applied to constant covariates.

At the initialization step (k=0) (see ***Figure 4—figure supplement 2***), the algorithm requests three inputs: (***World Health Organization, 2021***) a set of potential M covariates, labelled *Marker m* for $m \in \{1, \cdots, M\}$ (e.g., immunological markers); (***Cobey et al., 2021***) a set of P parameters on which covariates could be added, labelled θ_p for $p \in \{1, \cdots, P\}$(e.g., β and δ); and (***Kuzmina et al., 2021***) an initial model (e.g., the model without covariates), labelled M^0, with θ_p^0 being the definition of the parameter θ_p. At each step k>0, we note M^{k-1} the current model resulting in the model built in the step k−1. Then, each combination of markers and parameters that have not already been added in M^{k-1}, labelled $r\left(r \in \left\{Marker\, m \bigotimes \theta_p \notin M^{k-1} \mid m \in \{1, \ldots M\},\ p \in \{1, \ldots P\}\right\}\right)$, are considered and tested in an univariate manner (each relation r is independently added in M^{k-1} and ran). To this end, the parameter θ_p involved in this relationship r is modified as $\theta_p^k(t) = \theta_p^{k-1}(t) \times exp\left(\phi_m^p \times Marker_m(t)\right)$, where ϕ_m^p is the regression coefficient related to the marker and $Marker_m(t)$ being the trajectory of

the marker over time, while other parameters remain unchanged $\left(\forall\theta_q \notin r,\ \theta_q^k(t) = \theta_q^{k-1}(t)\right)$. Once all these models evaluated, the one with the optimal value of a given selection criterion defining the quality of the fits (e.g., the lowest BICc value) is selected and compared to the model M^{k-1}. If the value of the criterion is better than the one found for M^{k-1}, then this model is defined as the new current model, M^k, and the algorithm moves to the step k+1. Otherwise, the algorithm stops. The algorithm can also be stopped at the end of a fixed number of step k.

The objective of this algorithm being to identify mechanistic CoP, at each step, the selected model should respect, in addition to the best fits criterion, the two other criteria defining mCoP meaning the ability to capture the effect of the group of intervention and the ability to better explain the variability on individual parameters than the model adjusted for the group effect. To this end, we verify that in the selected model additionally adjusted for the group of intervention, the group effect appears as non-significantly different from 0 using a Wald test. Then, we check that the variances of random effects in the selected model are lower or equal to the ones obtained in the model adjusted only for the group effect.

Modelling hypothesis for time-dependent covariates in our application

Using a population-based approach to estimate our mechanistic model and similar to the adjustment of the model for constant covariates (e.g., groups of intervention), time-varying covariates are incorporated into the statistical model as individual-specific explanatory variables in the mixed-effects models. To implement the algorithm for selecting the time-varying covariates, many modelling choices were made. First, targeting covariates able to fully replace the group of intervention, we kept a similar mathematical relationship between parameters and immune markers than the one used with the constant covariate (see ***Equation 2***). Accordingly, we adjusted the model parameters additively in logarithmic scale. In this regard, at each step k (k>0), the parameter θ_p was defined as $\log\left(\theta_p^k(t)\right) = \log\left(\theta_p^{k-1}(t)\right) + \phi_m^p \times Marker_m(t)$. However, this choice may affect the results and other choices may be more relevant under different conditions. Second, because immune markers are observed only at discrete time points, whereas the estimation of the model is performed in a continuous way, we introduced immune markers as time-varying covariates using linear interpolation. Lets denote $Marker_{i,j}$ the value of the marker observed for the ith animal at the jth time point, with $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, J\}$. By linear interpolation, the time-continuous marker was defined as, $\forall t > 0$,

$$\boldsymbol{Marker_i^{int}}(\boldsymbol{t}) = \sum_{j=1}^{J-1} I_{[t_j;t_{j+1}]}(\boldsymbol{t})\left[\frac{\boldsymbol{Marker_{i,j+1}-Marker_{i,j}}}{\boldsymbol{t_{j+1}-t_j}}\boldsymbol{t} + \frac{\boldsymbol{Marker_{i,j}t_{j+1}-Marker_{j+1}t_j}}{\boldsymbol{t_{j+1}-t_j}}\right] + I_{\boldsymbol{t}\geq\boldsymbol{t_J}}(\boldsymbol{t}) \times \boldsymbol{Marker_{i,J}}$$

As previously described in the Results section, three different studies were considered in this work: a main study reported by ***Marlin et al., 2021***, testing the αCD40.RBD vaccine, and two additional studies (***Corbett et al., 2020***; ***Brouwer et al., 2021***) evaluating a two-component spike nanoparticle vaccine and the mRN-1273 vaccine, respectively. In the main study, the method was applied with both time-varying covariates and constant covariates for which only baseline value was considered, such that $Marker_i(t)=Marker_i(t=0)$ (see ***Supplementary file 1***). For the other two studies, only the baseline values were considered as covariates, the dynamics being not available. To assess the robustness of the results, several selection criteria were tested: AIC, BIC, log-likelihood, the percentage of explained interindividual variability, and similar results were obtained for all (results not shown). Moreover, as presented in Appendix 2 'BICc as selection criteria and multiple testing adjustment', we verified the robustness of the use of BIC as selection criteria despite the multiplicity of the tests. The identification of antibodies inhibiting the attachment of the RBD to the ACE2 receptor (ECLRBD) as the first time-varying CoP led to the definition of the time-varying viral infectivity for the ith animal as described in ***Equation 5***, while the selection anti-RBD IgG-binding antibodies led to the elimination rate of infected cells given in ***Equation 6***.

$$\beta_i(\boldsymbol{t}) = 10^{\boldsymbol{\beta_0}+\boldsymbol{u_i^\beta}} \times exp\left(\phi_{\boldsymbol{ecl}}^{\boldsymbol{\beta}} \times \boldsymbol{ECLRBD_i^{int}}(\boldsymbol{t})\right) \quad (5)$$

$$\boldsymbol{\delta_i}(\boldsymbol{t}) = \boldsymbol{\delta_0} \times \mathbf{exp}\left(\phi_{\boldsymbol{igg}}^{\boldsymbol{\delta}} \times \boldsymbol{IggRBD_i^{int}}(\boldsymbol{t}) + \boldsymbol{u_i^\delta}\right) \quad (6)$$

Quantification and statistical analysis

In each of the three studies used in this work, no statistical tests were performed on the raw data (i.e, observations), whether for viral load or for immune marker measurements, to identify statistical differences between treatment groups, as the statistical analyses were already been performed in the respective papers. Statistical significance of the effect of groups in model estimation is indicated in the tables by stars: *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$ and were estimated by Wald tests (Monolix software version 2019R1).

Model parameters were estimated with the SAEM algorithm (Monolix software version 2019R1). Graphics were generated using R version 3.6.1 and Excel 2016 and details on the statistical analysis for the experiments can be found in the accompanying figure legends. Horizontal red dashed lines on graphs indicate assay limit of detection.

Acknowledgements

We would like to thank J Guedj and O Terrier for fruitful discussions on the model definition. We thank S Langlois, J Demilly, N Dhooge, P Le Calvez, M Potier, JM Robert, T Prot, and C Dodan for the NHP experiments; L Bossevot, M Leonec, L Moenne-Loccoz, M Calpin-Lebreau, and J Morin for the RT-qPCR, ELISpot and Luminex assays, and for the preparation of reagents; A-S Gallouët, M Gomez-Pacheco, and W Gros for NHP T-cell assays and flow cytometry; B Fert for her help with the CT scans; M Barendji, J Dinh, and E Guyon for the NHP sample processing; S Keyser for the transports organization; F Ducancel and Y Gorin for their help with the logistics and safety management; I Mangeot for her help with resources management and B Targat contributed to data management. The monkey and syringe pictures in ***Figure 1*** was created with BioRender.com. This work was supported by INSERM and the Investissements d'Avenir program, Vaccine Research Institute (VRI), managed by the ANR under reference ANR-10-LABX-77-01. MA has been funded by INRIA PhD grant. The Infectious Disease Models and Innovative Therapies (IDMIT) research infrastructure is supported by the 'Programme Investissements d'Avenir', managed by the ANR under reference ANR-11-INBS-0008. The Fondation Bettencourt Schueller and the Region Ile-de-France contributed to the implementation of IDMIT's facilities and imaging technologies used to define volume of respiratory tract. The NHP study received financial support from REACTing, the Fondation pour la Recherche Medicale (FRM; AM-CoV-Path). We thank Lixoft SAS for their support. Numerical computations were in part carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP, and Conseil Régional d'Aquitaine (see https://www.plafrim.fr). We thank Miles Davenport and Frederik Graw as Senior Editor and Reviewing Editor of our paper, respectively, and the three anonymous reviewers for their time and their constructive comments.

Additional information

Funding

Funder	Grant reference number	Author
Agence Nationale de la Recherche	ANR-10-LABX-77-01	Yves Levy Rodolphe Thiébaut
Agence Nationale de la Recherche	ANR-11- 1018 INBS-0008	Roger Le Grand

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Marie Alexandre, Conceptualization, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review and editing; Romain Marlin, Conceptualization, Investigation, Resources, Visualization, Writing – original draft, Writing – review and editing; Mélanie Prague, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing; Severin Coleon, Sylvain Cardinaud, Benoit Delache, Mathieu Surenaud, Mathilde Galhaut,

Nathalie Dereuddre-Bosquet, Mariangela Cavarelli, Pauline Maisonnasse, Mireille Centlivre, Christine Lacabaratz, Aurelie Wiedemann, Sandra Zurawski, Gerard Zurawski, Investigation, Resources, Writing – review and editing; Nidhal Kahlaoui, Thibaut Naninck, Investigation, Resources, Visualization, Writing – review and editing; Olivier Schwartz, Rogier W Sanders, Resources, Writing – review and editing; Roger Le Grand, Yves Levy, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing; Rodolphe Thiébaut, Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review and editing

Author ORCIDs
Marie Alexandre http://orcid.org/0000-0002-3557-7075
Mélanie Prague http://orcid.org/0000-0001-9809-7848
Pauline Maisonnasse http://orcid.org/0000-0002-0555-207X
Olivier Schwartz http://orcid.org/0000-0002-0729-1475
Roger Le Grand http://orcid.org/0000-0002-4928-4484
Rodolphe Thiébaut http://orcid.org/0000-0002-5235-3962

Ethics
Cynomolgus macaques (Macaca fascicularis), aged 37–66 months (18 females and 13 males) and originating from Mauritian AAALAC certified breeding centers were used in this study. All animals were housed in IDMIT facilities (CEA, Fontenay-aux-roses), under BSL2 and BSL-3 containment when necessary (Animal facility authorization #D92-032-02, Préfecture des Hauts de Seine, France) and in compliance with European Directive 2010/63/EU, the French regulations and the Standards for Human Care and Use of Laboratory Animals, of the Office for Laboratory Animal Welfare (OLAW, assurance number #A5826-01, US). The protocols were approved by the institutional ethical committee "Comité d'Ethique en Expérimentation Animale du Commissariat à l'Energie Atomique et aux Energies Alternatives" (CEtEA #44) under statement number A20-011. The study was authorized by the "Research, Innovation and Education Ministry" under registration number APAFIS#24434–2020030216532863 v1.

Decision letter and Author response
Decision letter https://doi.org/10.7554/eLife.75427.sa1
Author response https://doi.org/10.7554/eLife.75427.sa2

Additional files

Supplementary files
- Supplementary file 1. Criteria defining neutralization as mechanistic correlate of protection of the effect of the vaccine on new cell infection.
- Supplementary file 2. Model parameters estimated by the model adjusted for groups of intervention.
- Supplementary file 3. Model parameters estimated by profile likelihood.
- Supplementary file 4. Model parameters estimated by the model adjusted for receptor-binding domain (RBD)/ACE2-binding inhibition on beta and for groups on delta.
- Transparent reporting form

Data availability
No unique reagents were generated for this study. Data that support the findings of this study are provided in the source data files of this paper and gather data from (1) the study [Marlin, Nature Com 2021] used in this analysis, which are also directly available online in the section Source data of this related paper (https://www.nature.com/articles/s41467-021-25382-0#Sec17); (2) the study [Brouwer, Cell 2021] used in this analysis, which are also available from the corresponding authors of the related paper and (3) the study [Corbett, NEJM 2020] used in this analysis, which are also available online in the section Supplementary Material of the related paper, excel file labelled ("Supplementary Appendix 2"). Data from the main study [Marlin, Nature Com 2021] can also be found in the open-access repository Dryad using the following DOI: https://doi.org/10.5061/dryad.1zcrjdfv7. The original code (mlxtran models and R) as well as model definition files including the full list of parameters used are available

and free-of-cost on github (Inria SISTM Team) at the following link: https://github.com/sistm/SARSCoV-2modelingNHP, (copy archived at swh:1:rev:a704c80daebc949434694d3f4441e48293c461cc).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Alexandre M, Marlin R, Prague M, Coleon S, Kahlaoui N, Cardinaud S, Naninck T, Delache B, Surenaud M, Galhaut M, Dereuddre-Bosquet N, Cavarelli M, Maisonnasse P, Centlivre M, Lacabaratz C, Wiedemann A, Zurawski S, Zurawski G, Schwartz O, Sanders RW, Le Grand R, Levy Y, Thiébaut R	2022	Viral loads and antibody, cytokine and T-cell responses in NHPs following vaccination targeting SARS-CoV-2 RBD domain to cells expressing CD40	https://doi.org/10.5061/dryad.1zcrjdfv7	Dryad Digital Repository, 10.5061/dryad.1zcrjdfv7

References

Alter G, Yu J, Liu J, Chandrashekar A, Borducchi EN, Tostanoski LH, McMahan K, Jacob-Dolan C, Martinez DR, Chang A, Anioke T, Lifton M, Nkolola J, Stephenson KE, Atyeo C, Shin S, Fields P, Kaplan I, Robins H, Amanat F, et al. 2021. Immunogenicity of Ad26.COV2.S vaccine against SARS-CoV-2 variants in humans. *Nature* **596**:268–272. DOI: https://doi.org/10.1038/s41586-021-03681-2, PMID: 34107529

Asgharian B, Price O, McClellan G, Corley R, Einstein DR, Jacob RE, Harkema J, Carey SA, Schelegle E, Hyde D, Kimbell JS, Miller FJ. 2012. Development of a rhesus monkey lung geometry model and application to particle deposition in comparison to humans. *Inhalation Toxicology* **24**:869–899. DOI: https://doi.org/10.3109/08958378.2012.725782, PMID: 23121298

Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS. 2006. Kinetics of influenza A virus infection in humans. *Journal of Virology* **80**:7590–7599. DOI: https://doi.org/10.1128/JVI.01623-05, PMID: 16840338

Balelli I, Pasin C, Prague M, Crauste F, Effelterre TV, Bockstal V, Solforosi L, Thiébaut R. 2020. A model for establishment, maintenance and reactivation of the immune response after vaccination against Ebola virus. *Journal of Theoretical Biology* **495**:110254. DOI: https://doi.org/10.1016/j.jtbi.2020.110254, PMID: 32205143

Bradfute SB, Bavari S. 2011. Correlates of immunity to filovirus infection. *Viruses* **3**:982–1000. DOI: https://doi.org/10.3390/v3070982, PMID: 21994766

Brouwer PJM, Brinkkemper M, Maisonnasse P, Dereuddre-Bosquet N, Grobben M, Claireaux M, de Gast M, Marlin R, Chesnais V, Diry S, Allen JD, Watanabe Y, Giezen JM, Kerster G, Turner HL, van der Straten K, van der Linden CA, Aldon Y, Naninck T, Bontjer I, et al. 2021. Two-component spike nanoparticle vaccine protects macaques from SARS-CoV-2 infection. *Cell* **184**:1188–1200. DOI: https://doi.org/10.1016/j.cell.2021.01.035, PMID: 33577765

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. 2006. Measurement Error in Nonlinear Models. Chapman and Hall/CRC. DOI: https://doi.org/10.1201/9781420010138

Chandrashekar A, Liu J, Martinot AJ, McMahan K, Mercado NB, Peter L, Tostanoski LH, Yu J, Maliga Z, Nekorchuk M, Busman-Sahay K, Terry M, Wrijil LM, Ducat S, Martinez DR, Atyeo C, Fischinger S, Burke JS, Slein MD, Pessaint L, et al. 2020. SARS-CoV-2 infection protects against rechallenge in rhesus macaques. *Science* **369**:812–817. DOI: https://doi.org/10.1126/science.abc4776

Charmet T, Schaeffer L, Grant R, Galmiche S, Chény O, Von Platen C, Maurizot A, Rogoff A, Omar F, David C, Septfons A, Cauchemez S, Gaymard A, Lina B, Lefrancois LH, Enouf V, van der Werf S, Mailles A, Levy-Bruhl D, Carrat F, et al. 2021. Impact of original, B.1.1.7, and B.1.351/P.1 SARS-CoV-2 lineages on vaccine effectiveness of two doses of COVID-19 mRNA vaccines: Results from a nationwide case-control study in France. *The Lancet Regional Health. Europe* **8**:100171. DOI: https://doi.org/10.1016/j.lanepe.2021.100171, PMID: 34278372

Cheng L, Wang Q, Li G, Banga R, Ma J, Yu H, Yasui F, Zhang Z, Pantaleo G, Perreau M, Zurawski S, Zurawski G, Levy Y, Su L. 2018. TLR3 agonist and CD40-targeting vaccination induces immune responses and reduces HIV-1 reservoirs. *The Journal of Clinical Investigation* **128**:4387–4396. DOI: https://doi.org/10.1172/JCI99005, PMID: 30148455

Chowdhury MZI, Turin TC. 2020. Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine and Community Health* **8**:e000262. DOI: https://doi.org/10.1136/fmch-2019-000262, PMID: 32148735

Chung AW, Kumar MP, Arnold KB, Yu WH, Schoen MK, Dunphy LJ, Suscovich TJ, Frahm N, Linde C, Mahan AE, Hoffner M, Streeck H, Ackerman ME, McElrath MJ, Schuitemaker H, Pau MG, Baden LR, Kim JH, Michael NL,

Barouch DH, et al. 2015. Dissecting Polyclonal Vaccine-Induced Humoral Immunity against HIV Using Systems Serology. *Cell* **163**:988–998. DOI: https://doi.org/10.1016/j.cell.2015.10.027, PMID: 26544943

Cobey S, Larremore DB, Grad YH, Lipsitch M. 2021. Concerns about SARS-CoV-2 evolution should not hold back efforts to expand vaccination. *Nature Reviews. Immunology* **21**:330–335. DOI: https://doi.org/10.1038/s41577-021-00544-9, PMID: 33795856

Corbett KS, Flynn B, Foulds KE, Francica JR, Boyoglu-Barnum S, Werner AP, Flach B, O'Connell S, Bock KW, Minai M, Nagata BM, Andersen H, Martinez DR, Noe AT, Douek N, Donaldson MM, Nji NN, Alvarado GS, Edwards DK, Flebbe DR, et al. 2020. Evaluation of the mRNA-1273 Vaccine against SARS-CoV-2 in Nonhuman Primates. *The New England Journal of Medicine* **383**:1544–1555. DOI: https://doi.org/10.1056/NEJMoa2024671

Dafni UG, Tsiatis AA. 1998. Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics* **54**:1445–1462. DOI: https://doi.org/10.2307/2533670, PMID: 9883544

Dagotto G, Yu J, Barouch DH. 2020. Approaches and Challenges in SARS-CoV-2 Vaccine Development. *Cell Host & Microbe* **28**:364–370. DOI: https://doi.org/10.1016/j.chom.2020.08.002, PMID: 32798444

Dufloo J, Grzelak L, Staropoli I, Madec Y, Tondeur L, Anna F, Pelleau S, Wiedemann A, Planchais C, Buchrieser J, Robinot R, Ungeheuer MN, Mouquet H, Charneau P, White M, Lévy Y, Hoen B, Fontanet A, Schwartz O, Bruel T. 2021. Asymptomatic and symptomatic SARS-CoV-2 infections elicit polyfunctional antibodies. *Cell Reports. Medicine* **2**:100275. DOI: https://doi.org/10.1016/j.xcrm.2021.100275, PMID: 33899033

Earle KA, Ambrosino DM, Fiore-Gartland A, Goldblatt D, Gilbert PB, Siber GR, Dull P, Plotkin SA. 2021. Evidence for antibody as a protective correlate for COVID-19 vaccines. *Vaccine* **39**:4423–4428. DOI: https://doi.org/10.1016/j.vaccine.2021.05.063, PMID: 34210573

Emary KRW, Golubchik T, Aley PK, Ariani CV, Angus B, Bibi S, Blane B, Bonsall D, Cicconi P, Charlton S, Clutterbuck EA, Collins AM, Cox T, Darton TC, Dold C, Douglas AD, Duncan CJA, Ewer KJ, Flaxman AL, Faust SN, et al. 2021. Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/01 (B.1.1.7): an exploratory analysis of a randomised controlled trial. *Lancet* **397**:1351–1362. DOI: https://doi.org/10.1016/S0140-6736(21)00628-0, PMID: 33798499

Eyal N, Lipsitch M. 2021. How to test SARS-CoV-2 vaccines ethically even after one is available. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* **73**:ciab182. DOI: https://doi.org/10.1093/cid/ciab182

Feng S, Phillips DJ, White T, Sayal H, Aley PK, Bibi S, Dold C, Fuskova M, Gilbert SC, Hirsch I, Humphries HE, Jepson B, Kelly EJ, Plested E, Shoemaker K, Thomas KM, Vekemans J, Villafana TL, Lambe T, Pollard AJ, et al. 2021. Correlates of protection against symptomatic and asymptomatic SARS-CoV-2 infection. *Nature Medicine* **27**:2032–2040. DOI: https://doi.org/10.1038/s41591-021-01540-1, PMID: 34588689

Fenwick C, Croxatto A, Coste AT, Pojer F, André C, Pellaton C, Farina A, Campos J, Hacker D, Lau K, Bosch B-J, Gonseth Nussle S, Bochud M, D'Acremont V, Trono D, Greub G, Pantaleo G. 2021. Changes in SARS-CoV-2 Spike versus Nucleoprotein Antibody Responses Impact the Estimates of Infections in Population-Based Seroprevalence Studies. *Journal of Virology* **95**:e01828-20. DOI: https://doi.org/10.1128/JVI.01828-20, PMID: 33144321

Flamar AL, Zurawski S, Scholz F, Gayet I, Ni L, Li XH, Klechevsky E, Quinn J, Oh S, Kaplan DH, Banchereau J, Zurawski G. 2012. Noncovalent assembly of anti-dendritic cell antibodies and antigens for evoking immune responses in vitro and in vivo. *Journal of Immunology* **189**:2645–2655. DOI: https://doi.org/10.4049/jimmunol.1102390, PMID: 22865916

Gaebler C, Wang Z, Lorenzi JCC, Muecksch F, Finkin S, Tokuyama M, Cho A, Jankovic M, Schaefer-Babajew D, Oliveira TY, Cipolla M, Viant C, Barnes CO, Bram Y, Breton G, Hägglöf T, Mendoza P, Hurley A, Turroja M, Gordon K, et al. 2021. Evolution of antibody immunity to SARS-CoV-2. *Nature* **591**:639–644. DOI: https://doi.org/10.1038/s41586-021-03207-w, PMID: 33461210

Godot V, Tcherakian C, Gil L, Cervera-Marzal I, Li G, Cheng L, Ortonne N, Lelièvre JD, Pantaleo G, Fenwick C, Centlivre M, Mouquet H, Cardinaud S, Zurawski SM, Zurawski G, Milpied P, Su L, Lévy Y. 2020. TLR-9 agonist and CD40-targeting vaccination induces HIV-1 envelope-specific B cells with a diversified immunoglobulin repertoire in humanized mice. *PLOS Pathogens* **16**:e1009025. DOI: https://doi.org/10.1371/journal.ppat.1009025, PMID: 33253297

Gonçalves A, Bertrand J, Ke R, Comets E, de Lamballerie X, Malvy D, Pizzorno A, Terrier O, Rosa Calatrava M, Mentré F, Smith P, Perelson AS, Guedj J. 2020. Timing of Antiviral Treatment Initiation is Critical to Reduce SARS-CoV-2 Viral Load. *CPT* **9**:509–514. DOI: https://doi.org/10.1002/psp4.12543, PMID: 32558354

Gonçalves A, Maisonnasse P, Donati F, Albert M, Behillil S, Contreras V, Naninck T, Marlin R, Solas C, Pizzorno A, Lemaitre J, Kahlaoui N, Terrier O, Ho Tsong Fang R, Enouf V, Dereuddre-Bosquet N, Brisebarre A, Touret F, Chapon C, Hoen B, et al. 2021. SARS-CoV-2 viral dynamics in non-human primates. *PLOS Computational Biology* **17**:e1008785. DOI: https://doi.org/10.1371/journal.pcbi.1008785, PMID: 33730053

Goyal A, Ribeiro RM, Perelson AS. 2017. The Role of Infected Cell Proliferation in the Clearance of Acute HBV Infection in Humans. *Viruses* **9**:E350. DOI: https://doi.org/10.3390/v9110350, PMID: 29156567

Goyal A, Liao LE, Perelson AS. 2019. Within-host mathematical models of hepatitis B virus infection: Past, present, and future. *Current Opinion in Systems Biology* **18**:27–35. DOI: https://doi.org/10.1016/j.coisb.2019.10.003, PMID: 31930181

Goyal A, Duke ER, Cardozo-Ojeda EF, Schiffer JT. 2020. Mathematical Modeling Explains Differential SARS CoV-2 Kinetics in Lung and Nasal Passages in Remdesivir Treated Rhesus Macaques. *bioRxiv*. DOI: https://doi.org/10.1101/2020.06.21.163550

Goyal A, Reeves DB, Schiffer JT. 2022. Multi-scale modelling reveals that early super-spreader events are a likely contributor to novel variant predominance. *Journal of the Royal Society, Interface* **19**:20210811. DOI: https://doi.org/10.1098/rsif.2021.0811, PMID: 35382576

Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. 2021. Thrombotic Thrombocytopenia after ChAdOx1 nCov-19 Vaccination. *The New England Journal of Medicine* **384**:2092–2101. DOI: https://doi.org/10.1056/NEJMoa2104840, PMID: 33835769

Gutmann C, Takov K, Burnap SA, Singh B, Ali H, Theofilatos K, Reed E, Hasman M, Nabeebaccus A, Fish M, McPhail MJ, O'Gallagher K, Schmidt LE, Cassel C, Rienks M, Yin X, Auzinger G, Napoli S, Mujib SF, Trovato F, et al. 2021. SARS-CoV-2 RNAemia and proteomic trajectories inform prognostication in COVID-19 patients admitted to intensive care. *Nature Communications* **12**:3406. DOI: https://doi.org/10.1038/s41467-021-23494-1, PMID: 34099652

Jin P, Li J, Pan H, Wu Y, Zhu F. 2021. Immunological surrogate endpoints of COVID-2019 vaccines: the evidence we have versus the evidence we need. *Signal Transduction and Targeted Therapy* **6**:48. DOI: https://doi.org/10.1038/s41392-021-00481-y, PMID: 33531462

Johnson M, Wagstaffe HR, Gilmour KC, Mai AL, Lewis J, Hunt A, Sirr J, Bengt C, Grandjean L, Goldblatt D. 2020. Evaluation of a novel multiplexed assay for determining IgG levels and functional activity to SARS-CoV-2. *Journal of Clinical Virology* **130**:104572. DOI: https://doi.org/10.1016/j.jcv.2020.104572, PMID: 32769024

Ke R, Zitzmann C, Ribeiro RM, Perelson AS. 2020. Kinetics of SARS-CoV-2 Infection in the Human Upper and Lower Respiratory Tracts and Their Relationship with Infectiousness. [medRxiv]. DOI: https://doi.org/10.1101/2020.09.25.20201772

Ke R, Zitzmann C, Ho DD, Ribeiro RM, Perelson AS. 2021. In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness. *PNAS* **118**:e2111477118. DOI: https://doi.org/10.1073/pnas.2111477118, PMID: 34857628

Kester KE, Cummings JF, Ofori-Anyinam O, Ockenhouse CF, Krzych U, Moris P, Schwenk R, Nielsen RA, Debebe Z, Pinelis E, Juompan L, Williams J, Dowler M, Stewart VA, Wirtz RA, Dubois M-C, Lievens M, Cohen J, Ballou WR, Heppner DG Jr, et al. 2009. Randomized, double-blind, phase 2a trial of falciparum malaria vaccines RTS,S/AS01B and RTS,S/AS02A in malaria-naive adults: safety, efficacy, and immunologic associates of protection. *The Journal of Infectious Diseases* **200**:337–346. DOI: https://doi.org/10.1086/600120, PMID: 19569965

Khoury DS, Cromer D, Reynaldi A, Schlub TE, Wheatley AK, Juno JA, Subbarao K, Kent SJ, Triccas JA, Davenport MP. 2021. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nature Medicine* **27**:1205–1211. DOI: https://doi.org/10.1038/s41591-021-01377-8, PMID: 34002089

Kim KS, Ejima K, Iwanami S, Fujita Y, Ohashi H, Koizumi Y, Asai Y, Nakaoka S, Watashi K, Aihara K, Thompson RN, Ke R, Perelson AS, Iwami S. 2021. A quantitative model used to compare within-host SARS-CoV-2, MERS-CoV, and SARS-CoV dynamics provides insights into the pathogenesis and treatment of SARS-CoV-2. *PLOS Biology* **19**:e3001128. DOI: https://doi.org/10.1371/journal.pbio.3001128, PMID: 33750978

Koch T, Mellinghoff SC, Shamsrizi P, Addo MM, Dahlke C. 2021. Correlates of Vaccine-Induced Protection against SARS-CoV-2. *Vaccines* **9**:238. DOI: https://doi.org/10.3390/vaccines9030238, PMID: 33801831

Kuzmina A, Khalaila Y, Voloshin O, Keren-Naus A, Boehm-Cohen L, Raviv Y, Shemer-Avni Y, Rosenberg E, Taube R. 2021. SARS-CoV-2 spike variants exhibit differential infectivity and neutralization resistance to convalescent or post-vaccination sera. *Cell Host & Microbe* **29**:522-528.. DOI: https://doi.org/10.1016/j.chom.2021.03.008, PMID: 33789085

Leung NHL, Chu DKW, Shiu EYC, Chan KH, McDevitt JJ, Hau BJP, Yen HL, Li Y, Ip DKM, Peiris JSM, Seto WH, Leung GM, Milton DK, Cowling BJ. 2020. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nature Medicine* **26**:676–680. DOI: https://doi.org/10.1038/s41591-020-0843-2, PMID: 32371934

Lui G, Ling L, Lai CK, Tso EY, Fung KS, Chan V, Ho TH, Luk F, Chen Z, Ng JK, Chow K-M, Cheng PK, Chan RC, Tsang DN, Gomersall CD, Hui DS, Chan PK. 2020. Viral dynamics of SARS-CoV-2 across a spectrum of disease severity in COVID-19. *The Journal of Infection* **81**:318–356. DOI: https://doi.org/10.1016/j.jinf.2020.04.014, PMID: 32315724

Lustig Y, Nemet I, Kliker L, Zuckerman N, Yishai R, Alroy-Preis S, Mendelson E, Mandelboim M. 2021. Neutralizing Response against Variants after SARS-CoV-2 Infection and One Dose of BNT162b2. *The New England Journal of Medicine* **384**:2453–2454. DOI: https://doi.org/10.1056/NEJMc2104036, PMID: 33826815

Marc A, Kerioui M, Blanquart F, Bertrand J, Mitjà O, Corbacho-Monné M, Marks M, Guedj J. 2021. Quantifying the relationship between SARS-CoV-2 viral load and infectiousness. *eLife* **10**:e69302. DOI: https://doi.org/10.7554/eLife.69302, PMID: 34569939

Marks M, Millat-Martinez P, Ouchi D, Roberts CH, Alemany A, Corbacho-Monné M, Ubals M, Tobias A, Tebé C, Ballana E, Bassat Q, Baro B, Vall-Mayans M, G-Beiras C, Prat N, Ara J, Clotet B, Mitjà O. 2021. Transmission of COVID-19 in 282 clusters in Catalonia, Spain: a cohort study. *The Lancet. Infectious Diseases* **21**:629–636. DOI: https://doi.org/10.1016/S1473-3099(20)30985-3, PMID: 33545090

Marlin R, Godot V, Cardinaud S, Galhaut M, Coleon S, Zurawski S, Dereuddre-Bosquet N, Cavarelli M, Gallouët AS, Maisonnasse P, Dupaty L, Fenwick C, Naninck T, Lemaitre J, Gomez-Pacheco M, Kahlaoui N, Contreras V, Relouzat F, Fang RHT, Wang Z, et al. 2021. Targeting SARS-CoV-2 receptor-binding domain to cells expressing CD40 improves protection to infection in convalescent macaques. *Nature Communications* **12**:5215. DOI: https://doi.org/10.1038/s41467-021-25382-0, PMID: 34471122

McMahan K, Yu J, Mercado NB, Loos C, Tostanoski LH, Chandrashekar A, Liu J, Peter L, Atyeo C, Zhu A, Bondzie EA, Dagotto G, Gebre MS, Jacob-Dolan C, Li Z, Nampanya F, Patel S, Pessaint L, Van Ry A, Blade K,

et al. 2021. Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature* **590**:630–634. DOI: https://doi.org/10.1038/s41586-020-03041-6, PMID: 33276369

Mercado NB, Zahn R, Wegmann F, Loos C, Chandrashekar A, Yu J, Liu J, Peter L, McMahan K, Tostanoski LH, He X, Martinez DR, Rutten L, Bos R, Vellinga J, Custers J, Langedijk JP, Kwaks T, Bakkers MJG, Zuijdgeest D, et al. 2020. Single-shot Ad26 vaccine protects against SARS-CoV-2 in rhesus macaques. *Nature* **586**:583–588. DOI: https://doi.org/10.1038/s41586-020-2607-z

Miao H, Xia X, Perelson AS, Wu H. 2011. ON IDENTIFIABILITY OF NONLINEAR ODE MODELS AND APPLICATIONS IN VIRAL DYNAMICS. *SIAM Review. Society for Industrial and Applied Mathematics* **53**:3–39. DOI: https://doi.org/10.1137/090757009, PMID: 21785515

Muñoz-Fontela C. 2020. Animal models for COVID-19. *Nature* **586**:509–515. DOI: https://doi.org/10.1038/s41586-020-2787-6

Myers MA, Smith AP, Lane LC, Moquin DJ, Aogo R, Woolard S, Thomas P, Vogel P, Smith AM. 2021. Dynamically linking influenza virus infection kinetics, lung injury, inflammation, and disease severity. *eLife* **10**:e68864. DOI: https://doi.org/10.7554/eLife.68864, PMID: 34282728

Néant N. 2021. Groups, Modeling SARS-CoV-2 viral kinetics and association with mortality in hospitalized patients from the French COVID cohort. *PNAS* **118**. DOI: https://doi.org/10.1073/pnas.2017962118

Nyberg T, Ferguson NM, Nash SG, Webster HH, Flaxman S, Andrews N, Hinsley W, Bernal JL, Kall M, Bhatt S, Blomquist P, Zaidi A, Volz E, Aziz NA, Harman K, Funk S, Abbott S, Hope R, Charlett A, Chand M, et al. 2022. Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B.1.1.529) and delta (B.1.617.2) variants in England: a cohort study. *Lancet* **399**:1303–1312. DOI: https://doi.org/10.1016/S0140-6736(22)00462-7, PMID: 35305296

Pinky L, Burke CW, Russell CJ, Smith AM. 2021. Quantifying dose-, strain-, and tissue-specific kinetics of parainfluenza virus infection. *PLOS Computational Biology* **17**:e1009299. DOI: https://doi.org/10.1371/journal.pcbi.1009299, PMID: 34383757

Pizzorno A, Padey B, Julien T, Trouillet-Assant S, Traversier A, Errazuriz-Cerda E, Fouret J, Dubois J, Gaymard A, Lescure FX, Dulière V, Brun P, Constant S, Poissy J, Lina B, Yazdanpanah Y, Terrier O, Rosa-Calatrava M. 2020. Characterization and Treatment of SARS-CoV-2 in Nasal and Bronchial Human Airway Epithelia. *Cell Reports. Medicine* **1**:100059. DOI: https://doi.org/10.1016/j.xcrm.2020.100059, PMID: 32835306

Planas D, Bruel T, Grzelak L, Guivel-Benhassine F, Staropoli I, Porrot F, Planchais C, Buchrieser J, Rajah MM, Bishop E, Albert M, Donati F, Prot M, Behillil S, Enouf V, Maquart M, Smati-Lafarge M, Varon E, Schortgen F, Yahyaoui L, et al. 2021. Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. *Nature Medicine* **27**:917–924. DOI: https://doi.org/10.1038/s41591-021-01318-5, PMID: 33772244

Plotkin SA, Gilbert PB. 2012. Nomenclature for immune correlates of protection after vaccination. *Clinical Infectious Diseases* **54**:1615–1617. DOI: https://doi.org/10.1093/cid/cis238, PMID: 22437237

Plotkin SA. 2013. Complex correlates of protection after vaccination. *Clinical Infectious Diseases* **56**:1458–1465. DOI: https://doi.org/10.1093/cid/cit048, PMID: 23386629

Plotkin SA. 2020. Updates on immunologic correlates of vaccine-induced protection. *Vaccine* **38**:2250–2257. DOI: https://doi.org/10.1016/j.vaccine.2019.10.046, PMID: 31767462

Pollard AJ, Bijker EM. 2021. A guide to vaccinology: from basic principles to new developments. *Nature Reviews. Immunology* **21**:83–100. DOI: https://doi.org/10.1038/s41577-020-00479-7, PMID: 33353987

Rhodes SJ, Guedj J, Fletcher HA, Lindenstrøm T, Scriba TJ, Evans TG, Knight GM, White RG. 2018. Using vaccine Immunostimulation/Immunodynamic modelling methods to inform vaccine dose decision-making. *NPJ Vaccines* **3**:36. DOI: https://doi.org/10.1038/s41541-018-0075-3, PMID: 30245860

Robinot R, Hubert M, de Melo GD, Lazarini F, Bruel T, Smith N, Levallois S, Larrous F, Fernandes J, Gellenoncourt S, Rigaud S, Gorgette O, Thouvenot C, Trébeau C, Mallet A, Duménil G, Gobaa S, Etournay R, Lledo P-M, Lecuit M, et al. 2021. SARS-CoV-2 infection induces the dedifferentiation of multiciliated cells and impairs mucociliary clearance. *Nature Communications* **12**:4354. DOI: https://doi.org/10.1038/s41467-021-24521-x, PMID: 34272374

Sawicki SG, Sawicki DL, Siddell SG. 2007. A contemporary view of coronavirus transcription. *Journal of Virology* **81**:20–29. DOI: https://doi.org/10.1128/JVI.01358-06, PMID: 16928755

Smith AP, Moquin DJ, Bernhauerova V, Smith AM. 2018. Influenza Virus Infection Model With Density Dependence Supports Biphasic Viral Decay. *Frontiers in Microbiology* **9**:1554. DOI: https://doi.org/10.3389/fmicb.2018.01554, PMID: 30042759

Tauzin A, Nayrac M, Benlarbi M, Gong SY, Gasser R, Beaudoin-Bussières G, Brassard N, Laumaea A, Vézina D, Prévost J, Anand SP, Bourassa C, Gendron-Lepage G, Medjahed H, Goyette G, Niessl J, Tastet O, Gokool L, Morrisseau C, Arlotto P, et al. 2021. A single dose of the SARS-CoV-2 vaccine BNT162b2 elicits Fc-mediated antibody effector functions and T cell responses. *Cell Host & Microbe* **29**:1137-1150.. DOI: https://doi.org/10.1016/j.chom.2021.06.001, PMID: 34133950

Thai HT, Mentré F, Holford NHG, Veyrat-Follet C, Comets E. 2014. Evaluation of bootstrap methods for estimating uncertainty of parameters in nonlinear mixed-effects models: a simulation study in population pharmacokinetics. *Journal of Pharmacokinetics and Pharmacodynamics* **41**:15–33. DOI: https://doi.org/10.1007/s10928-013-9343-z, PMID: 24317870

Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, Chang S, Conley SD, Mori Y, Seita J, Berry GJ, Shrager JB, Metzger RJ, Kuo CS, Neff N, Weissman IL, Quake SR, Krasnow MA. 2020. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**:619–625. DOI: https://doi.org/10.1038/s41586-020-2922-4, PMID: 33208946

Vanshylla K, Di Cristanziano V, Kleipass F, Dewald F, Schommers P, Gieselmann L, Gruell H, Schlotz M, Ercanoglu MS, Stumpf R, Mayer P, Zehner M, Heger E, Johannis W, Horn C, Suárez I, Jung N, Salomon S, Eberhardt KA, Gathof B, et al. 2021. Kinetics and correlates of the neutralizing antibody response to SARS-CoV-2 infection in humans. *Cell Host & Microbe* **29**:917-929.. DOI: https://doi.org/10.1016/j.chom.2021.04.015, PMID: 33984285

Wang S, Pan Y, Wang Q, Miao H, Brown AN, Rong L. 2020. Modeling the viral dynamics of SARS-CoV-2 infection. *Mathematical Biosciences* **328**:108438. DOI: https://doi.org/10.1016/j.mbs.2020.108438, PMID: 32771304

World Health Organization. 2021. World Health Organisation, COVID-19 vaccine tracker and landscape. COVID-19. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines)

Wu L. 2009. Mixed Effects Models for Complex Data. Chapman and Hall/CRC. DOI: https://doi.org/10.1201/9781420074086

Yu J, Tostanoski LH, Peter L, Mercado NB, McMahan K, Mahrokhian SH, Nkolola JP, Liu J, Li Z, Chandrashekar A, Martinez DR, Loos C, Atyeo C, Fischinger S, Burke JS, Slein MD, Chen Y, Zuiani A, Lelis FJN, Travers M, et al. 2020. DNA vaccine protection against SARS-CoV-2 in rhesus macaques. *Science* **369**:806–811. DOI: https://doi.org/10.1126/science.abc6284, PMID: 32434945

Zhou D, Dejnirattisai W, Supasa P, Liu C, Mentzer AJ, Ginn HM, Zhao Y, Duyvesteyn HME, Tuekprakhon A, Nutalai R, Wang B, Paesen GC, Lopez-Camacho C, Slon-Campos J, Hallis B, Coombes N, Bewley K, Charlton S, Walter TS, Skelly D, et al. 2021. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184**:2348-2361.. DOI: https://doi.org/10.1016/j.cell.2021.02.037, PMID: 33730597

Zost SJ, Gilchuk P, Chen RE, Case JB, Reidy JX, Trivette A, Nargi RS, Sutton RE, Suryadevara N, Chen EC, Binshtein E, Shrihari S, Ostrowski M, Chu HY, Didier JE, MacRenaris KW, Jones T, Day S, Myers L, Eun-Hyung Lee F, et al. 2020. Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nature Medicine* **26**:1422–1427. DOI: https://doi.org/10.1038/s41591-020-0998-x, PMID: 32651581

Zurawski G, Shen X, Zurawski S, Tomaras GD, Montefiori DC, Roederer M, Ferrari G, Lacabaratz C, Klucar P, Wang Z, Foulds KE, Kao SF, Yu X, Sato A, Yates NL, LaBranche C, Stanfield-Oakley S, Kibler K, Jacobs B, Salazar A, et al. 2017. Superiority in Rhesus Macaques of Targeting HIV-1 Env gp140 to CD40 versus LOX-1 in Combination with Replication-Competent NYVAC-KC for Induction of Env-Specific Antibody and T Cell Responses. *Journal of Virology* **91**:e01596-16. DOI: https://doi.org/10.1128/JVI.01596-16, PMID: 28202751

Appendix 1

Model building

In the model presented in the manuscript, we considered the two compartments of the URT, trachea, and nasopharynx, as two distinct compartments (i.e., without transfer of virus between them), as described by ***Equation AE1***. In each of them, the viral dynamics are described by a target-cell limited model augmented with a compartment describing the dynamics of the inoculated virus (V_s). Moreover, in the statistical model describing the model parameters, the three parameters β, δ, and P were assumed as jointly estimated between the two compartments, with shared random effects and covariates and considering that parameters β and δ are equal in both trachea and nasopharynx ($\beta^T = \beta^N$, $\delta^T = \delta^N$).

$$\begin{cases} \frac{dT_i^N}{dt} = -\beta^N V_i^N T^N - \mu\beta^N V_s^N T^N \\ \frac{dI_1^N}{dt} = \beta^N V_i^N T^N + \mu\beta^N V_s^N T^N - kI_1^N \\ \frac{dI_2^N}{dt} = kI_1^N - \delta^N I_2^N \\ \frac{dV_i^N}{dt} = \mu P^N I_2^N - cV_i^N - \beta^N V_i^N T^N \\ \frac{dV_{ni}^N}{dt} = (1-\mu) P^N I_2^N - cV_{ni}^N \\ \frac{dV_s^N}{dt} = -c_i V_s^N - \mu\beta^N V_s^N T^N \end{cases} \quad \begin{cases} \frac{dT_i^T}{dt} = -\beta^T V_i^T T^T - \mu\beta^T V_s^T T^T \\ \frac{dI_1^T}{dt} = \beta^T V_i^T T^T + \mu\beta^T V_s^T T^T - kI_1^T \\ \frac{dI_2^T}{dt} = kI_1^T - \delta^T I_2^T \\ \frac{dV_i^T}{dt} = \mu P^T I_2^T - cV_i^T - \beta^T V_i^T T^T \\ \frac{dV_{ni}^T}{dt} = (1-\mu) P^T I_2^T - cV_{ni}^T \\ \frac{dV_s^T}{dt} = -c_i V_s^T - \mu\beta^T V_s^T T^T \end{cases} \quad \text{(AE1)}$$

Initially, random effects were added on the three parameters. However, taken into consideration identifiability issues that are usually encountered between the viral infectivity (β) and the viral production (P), we decided to remove the possibility of interindividual variability on the parameter P. This choice was also driven by multiple model estimations showing less robust estimations when variability was allowed in both parameters β and P. In particular, the estimate of the viral production was impacted by a ratio between the parameter and its standard error (RSE) higher than 100%.

Comparison of the parameters between the tracheal and the nasopharyngeal compartments

To decide which of these three parameters were assumed to be equal between the two compartments, all possibilities were tested and compared, using the BICc as selection criteria. As shown in ***Appendix 1—table 1***, we started with the model in which all parameters were equal between the two compartments and we progressively relaxed this hypothesis. During this step, no exchange of virions between the two compartments of the URT was possible (g=0). Once all models estimated, we kept the one with the lowest value of BICc, meaning with the highest negative difference of BICc compared to the initial model. We identified the model with only the viral production varying between the two compartments as the best one to fit the data.

Appendix 1—table 1. Comparison of models evaluating the difference of viral infectivity (β), loss of infected cells (δ), and viral production (P) between the nasopharynx and the trachea.

Model tested	Statistical model	ΔBICc
Initial model	$\beta^T = \beta^N$ $\delta^N = \delta^T$ $P^N = P^T$ Variability on β and δ	
Model with different β	$\boldsymbol{\beta^T \neq \beta^N}$ $\delta^N = \delta^T$ $P^N = P^T$ Variability on β and δ	–17.31
Model with different δ	$\beta^T = \beta^N$ $\boldsymbol{\delta^N \neq \delta^T}$ $P^N = P^T$ Variability on β and δ	–14.38

Appendix 1—table 1 Continued on next page

Appendix 1—table 1 Continued

Model tested	Statistical model	ΔBICc
Model with different P	$\beta^T = \beta^N$ $\delta^N = \delta^T$ $P^N \neq P^T$ Variability on β and δ	**–25.24**
Model with different β and δ	$\beta^T \neq \beta^N$ $\delta^N \neq \delta^T$ $P^N = P^T$ Variability on β and δ	–13.00
Model with different β and P	$\beta^T \neq \beta^N$ $\delta^N = \delta^T$ $P^N \neq P^T$ Variability on β and δ	–19.19
Model with different δ and P	$\beta^T = \beta^N$ $\delta^N \neq \delta^T$ $P^N \neq P^T$ Variability on β and δ	–19.47
Model with different β, δ, and P	$\beta^T \neq \beta^N$ $\delta^N \neq \delta^T$ $P^N \neq P^T$ Variability on β and δ	–13.39

Identification of group effects

Once the structure of the statistical model defined, we tried to identify on which parameters an effect of the group of treatment could be identified and by extension on which biological mechanisms. In this step, we were interested in four parameters: β, δ, P, and c, the latter being the clearance of de novo- produced virions. In the study, three groups of treatments were considered as constant categorical covariates: naïve, convalescent, and convalescent vaccinated. We performed a forward selection approach using the BICc as selection criteria to find the best model, using the model without covariate as initial model. At each step the model decreasing the most the value of the BICc is selected and the procedure stops once the BICc does not decrease anymore. At each step of the procedure, the statistical significance of covariate added into the model was verified via a Wald test. As shown in ***Appendix 1—table 2***, the selected model identified a group effect on the viral infectivity and the loss rate of infected cells.

Appendix 1—table 2. Comparison of models evaluating the adjustment of the viral infectivity (β), the loss rate of infected cells (δ), the viral production (P), and the viral clearance (c) for the groups of treatment.

The group of naïve animals is assumed as the group of reference.

Step	Model tested	Statistical model	$\Delta BICc$
	Initial model: Model without group effects	$\beta = 10^{\beta_0}$ $\delta = \delta_0$ $P = P_0$ $c = c_0$	
	Model with group effect on β	$\beta = 10^{\left(\beta_0 + \phi^{\beta}_{conv} + \phi^{\beta}_{CD40}\right)}$ $\delta = \delta_0$ $P = P_0$ $c = c_0$	–21.5
	Model with group effect on δ	$\beta = 10^{\beta_0}$ $\delta = \delta_0 \exp\left(\phi^{\delta}_{conv} + \phi^{\delta}_{CD40}\right)$ $P = P_0$ $c = c_0$	–16.62
	Model with group effect on P	$\beta = 10^{\beta_0}$ $\delta = \delta_0$ $P = P_0 \exp\left(\phi^{P}_{conv} + \phi^{P}_{CD40}\right)$ $c = c_0$	+9.68
1	Model with group effect on c	$\beta = 10^{\beta_0}$ $\delta = \delta_0$ $P = P_0$ $c = c_0 \exp\left(\phi^{c}_{conv} + \phi^{c}_{CD40}\right)$	+9.20
	Initial model: Model with group effect on β	$\beta = 10^{\left(\beta_0 + \phi^{\beta}_{conv} + \phi^{\beta}_{CD40}\right)}$ $\delta = \delta_0$ $P = P_0$ $c = c_0$	
	Model with group effect on β and δ	$\boldsymbol{\beta = 10^{\left(\beta_0 + \phi^{\beta}_{conv} + \phi^{\beta}_{CD40}\right)}}$ $\boldsymbol{\delta = \delta_0 \exp\left(\phi^{\delta}_{conv} + \phi^{\delta}_{CD40}\right)}$ $P = P_0$ $c = c_0$	–2.48
	Model with group effect on β and P	$\beta = 10^{\left(\beta_0 + \phi^{\beta}_{conv} + \phi^{\beta}_{CD40}\right)}$ $\delta = \delta_0$ $P = P_0 \exp\left(\phi^{P}_{conv} + \phi^{P}_{CD40}\right)$ $c = c_0$	+12.25
2	Model with group effect on β and c	$\beta = 10^{\left(\beta_0 + \phi^{\beta}_{conv} + \phi^{\beta}_{CD40}\right)}$ $\delta = \delta_0$ $P = P_0$ $c = c_0 \exp\left(\phi^{c}_{conv} + \phi^{c}_{CD40}\right)$	+11.97
	Initial model: Model with group effect on β and δ	$\beta = 10^{\left(\beta_0 + \phi^{\beta}_{conv} + \phi^{\beta}_{CD40}\right)}$ $\delta = \delta_0 \exp\left(\phi^{\delta}_{conv} + \phi^{\delta}_{CD40}\right)$ $P = P_0$ $c = c_0$	
	Model with group effect on β, δ, and P	$\beta = 10^{\left(\beta_0 + \phi^{\beta}_{conv} + \phi^{\beta}_{CD40}\right)}$ $\delta = \delta_0 \exp\left(\phi^{\delta}_{conv} + \phi^{\delta}_{CD40}\right)$ $P = P_0 \exp\left(\phi^{P}_{conv} + \phi^{P}_{CD40}\right)$ $c = c_0$	+10.88
3	Model with group effect on β, δ, and c	$\beta = 10^{\left(\beta_0 + \phi^{\beta}_{conv} + \phi^{\beta}_{CD40}\right)}$ $\delta = \delta_0 \exp\left(\phi^{\delta}_{conv} + \phi^{\delta}_{CD40}\right)$ $P = P_0$ $c = c_0 \exp\left(\phi^{c}_{conv} + \phi^{c}_{CD40}\right)$	+11.61

Based on all these results, the optimal statistical model with adjustment for groups of treatment was defined as follows:

$$\begin{cases} \log_{10}\left(\beta_i\right) = \beta_0 + \phi^{\beta}_{conv} \times I_{i\in conv} + \phi^{\beta}_{CD40} \times I_{i\in CD40} + u^{\beta}_i \\ \log\left(\delta_i\right) = \log\left(\delta_0\right) + \phi^{\delta}_{conv} \times I_{i\in conv} + \phi^{\delta}_{CD40} \times I_{i\in CD40} + u^{\delta}_i \\ \log\left(P^N_i\right) = \log\left(P_0\right) \\ P^T_i = P^N_i \times \exp\left(f^T_P\right) \end{cases}$$

Exchange of viruses between the nasopharyngeal and tracheal compartments

Afterward, we tested the possibility of an exchange of free plasma virus from between the two compartments of the URT. We made the hypothesis of a constant first-order exchange and we tested the addition a transfer of virions from nasopharyngeal to tracheal compartments and vice versa, with a migration rate g_{NT} and g_{TN}, respectively. To this end, equations of infectious (V_i) and non-infectious (V_{ni}) viruses in ***Equation AE1*** between the two compartments were linked as follows:

$$\begin{array}{ll} \frac{dV^T_i}{dt} \mapsto \frac{dV^T_i}{dt} - g_{TN}V^T_i + g_{NT}V^N_i & \frac{dV^T_{ni}}{dt} \mapsto \frac{dV^T_{ni}}{dt} - g_{TN}V^T_{ni} + g_{NT}V^N_{ni} \\ \frac{dV^N_i}{dt} \mapsto \frac{dV^N_i}{dt} + g_{TN}V^T_i - g_{NT}V^N_i & \frac{dV^N_{ni}}{dt} \mapsto \frac{dV^N_{ni}}{dt} + g_{TN}V^T_{ni} - g_{NT}V^N_{ni} \end{array} \quad \text{(AE2)}$$

with the arrow symbolizing the modification of the equations defined in ***Equation AE1*** and g_{NT} and g_{TN} being two positive rates. As a first step, we tried to estimate either bidirectional or one of the two unidirectional transfers using the data from the 18 NHPs of the first study described in the main paper. However, data were too spare to bring enough information to get estimations. Consequently, as a second step, additional data were used: two naïve macaques were exposed to the same dose (1×10^6 pfu) of SARS-CoV-2 than the 18 NHPs of the main study. However, instead of being inoculated via intra-tracheal (4.5 mL) and intra-nasal (0.5 mL) routes, these latter received inoculum via intra-gastric (4.5 mL) and intra-nasal (0.5 mL) routes. Similar to the main study, the viral gRNA dynamics in both tracheal and nasopharyngeal compartments were repeatedly measured during the 20 days following the challenge (***Figure 2—figure supplement 2E***).

These two additional macaques having not received intra-tracheal inoculum, viral dynamics measured in this same compartment was expected to come from (at least partially) an exchange with the nasopharynx and thus bring information about it. However, having only two macaques without virions inoculated via intra-tracheal route, no enough information were available to totally estimate the model with exchanges. Consequently, these two additional NHPs having similar characteristics than the 18 NHPs involved in the main study, we made the assumption that the viral dynamics in nasopharynx after inoculation and the viral dynamics in the trachea, once the transfer initiated, should be described by the same model (without inoculum in trachea) and those by the same parameters. We expected that the difference of dynamics in trachea between these two set of macaques could allow an estimation of the parameters g_{TN} and/or g_{NT}. For that reason, we estimated the model in ***Equation AE1*** using data from the 18 NHPs of the main study. Then using the data from the two additional NHPs, and assuming all parameters of the model resulting from ***Equation AE2*** as fixed (see ***Appendix 1—table 2***), except g_{TN} and g_{NT}, we tried to quantify the transfers of virions.

The estimation of multiple models on those two animals tended to conclude that only a unidirectional transfer of viruses from the nasopharyngeal to the tracheal compartment should be explored, with an estimation of g_{NT} ranging from 0.9 to 2.5 day^{-1}. Once these values quantified, we tried to update/re-estimate the model, initially estimated on the 18 NHPs, using only a unidirectional transfer from nasopharynx to trachea and fixing the value of the migration rate at the different values aforementioned. However, all tested values of g_{NT} led irremediably to a degradation of the model with an increase of at least two points of BICc.

An estimation of the parameter g_{NT} by profile likelihood (results not shown) led to a strictly increasing profile of the likelihood (the lower the better) and was thus no more conclusive. Consequently, no exchange of virions were assumed in the final model and the parameters g_{NT} and g_{TN} were fixed at 0 day^{-1}.

Appendix 2

BICc as selection criteria and multiple testing adjustment

In the case of classic covariate selection approaches using p-values as selection criteria, particular attention must be paid to take into account the dependence of the results on the number tests performed.

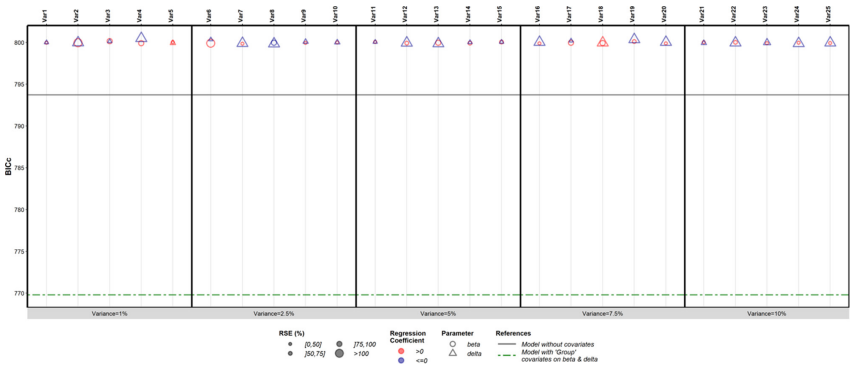
Over the years, multiple corrections have been proposed to adjust results for test multiplicity (e.g., Bonferroni correction, Benjamini and Hochberg correction among others).

Although we verified the significance of the covariate selected in our model, our covariate selection approach relies on the BICc. To ensure the robustness of the BICc as selection criterion despite the multiplicity of the tests, we performed an additional simulation work.

We simulated M=25 longitudinal variables for 18 individuals and with similar time points than those found on our data, meaning at days 0, 4, 9, and 20 post-infection. Variables were simulated as white-noise random variables such that for the ith subject at the jth time point, the mth variable was defined as $X_{ij}^{m} \sim \mathfrak{N}\left(0,\ \sigma^2\right)$, with m=1, …, M. In our simulations, we tested five values for the variance σ^2 ranging from 1% to 10% (five variables simulated for each value of σ).

Assuming these variables as our time-varying covariates, we applied the forward selection approach used in our method by testing each of them in a univariate manner of both β and δ.

As shown in ***Appendix 2—figure 1***, the 50 models built to evaluate the adjustment of either β or δ for the simulated variables provide similar results in terms of BICc, and thus whatever the value of the standard deviation σ used. Consequently, these results appear as quite robust to the multiplicity of the test. Moreover, as expected, adjustments for white-noise random variables depict the degradation of the model in comparison to the model without covariates.



Appendix 2—figure 1. Results of the forward selection approach applied on the 25 simulated white-noise random variables. The discrete *x*-axis represents the different variables and the *y*-axis represents the values of the corrected Bayesian information criteria (BICc). Circles and triangles correspond to the results obtained with the parameters β or δ adjusted for the variables. The horizontal solid black line represents the value of the BICc obtained with the model without covariates while the horizontal dashed green line highlights the value of the criterion obtained with both β and δ adjusted for the groups of treatment.

Evaluation of the robustness of the estimation

To evaluate the robustness of the parameter estimates obtained on our models, despite the small number of independent observations, we performed a bootstrap procedure with replacement (***Thai et al., 2014***), for B=50 iterations. The bootstrap parameter estimate was calculated as the median of the parameter estimates from the B bootstrap samples while the standard error of each parameter was calculated according to the definition of ***Thai et al., 2014***, which means with the SE of the lth component of the vector of parameters given by:

$$\hat{SE}_B^J = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}_b^{*(l)} - \hat{\theta}_B^{(l)}\right)^2}$$

with $\theta_b^{*(l)}$ being its estimate obtained at the bth iteration of the bootstrap and $\theta_B^{(l)}$ the bootstrap parameter estimate. For each bootstrap sample, we paid attention to keep the 1:1:1 ratio between the three groups of treatment, with six animals selected within each group. Results are reported in ***Appendix 2—table 1*** and ***Appendix 2—table 2***.

Appendix 2—table 1. Model parameters for viral dynamics in both the nasopharynx and the trachea estimated by the model adjusted for groups of intervention.
For the bootstrap procedure, 50 iterations were performed.

Parameter	Meaning	Value [95% CI]	Unit
β	Viral infectivity in the naive group ($\times 10^{-6}$)	0.91 [0.12; 7.03]	(copies/mL)$^{-1}$ day^{-1}
	Fold change in the convalescent group	0.15 [0.04; 0.58]	
	Fold change in the Conv-CD40 group	0.006 [0.001; 0.04]	
δ	Loss rate of infected cells in the naive group	1.09 [0.74; 1.60]	day^{-1}
	Fold change in the convalescent group	1.70 [1.08; 2.66]	
	Fold change in the Conv-CD40 group	2.00 [0.94; 4.27]	
P^N	Viral production rate in the naso. ($\times 10^3$)	10.1 [1.16; 87.7]	virions (cell day)$^{-1}$
P^T	Viral production rate in the trachea ($\times 10^3$)	0.86 [0.08; 9.19]	virions (cell day)$^{-1}$
$\alpha_{v/sg}$	Infected cells and sgRNA viral load ratio	1.42 [0.99; 2.02]	virions cell^{-1}
k	Eclipse rate	3	day^{-1}
c	Clearance of de novo produced viruses	3	day^{-1}
c_I	Clearance of inoculum	20	day^{-1}
μ	Percentage of infectious viruses	10^{-3}	
$T_0^{X,nbc}$	Initial number of target cells	1.25×10^5 (naso.) 2.25×10^4 (trachea)	cells
$Inoc_0$	Number of virions inoculated	2.19×10^{10}	virions
ω_β	SD of random effect on $\log_{10}\beta$	0.319 [0.111; 0.527]	
ω_δ	SD of random effect on δ	0.122 [-0.039; 0.283]	
σ_{VLn}	SD of error model gRNA in naso.	1.24 [0.96; 1.51]	
σ_{VLt}	SD of error model gRNA in trachea	1.09 [0.92; 1.26]	
σ_{sgVLn}	SD of error model sgRNA in naso	1.35 [1.08; 1.61]	
σ_{sgVLt}	SD of error model sgRNA in trachea	1.53 [1.15; 1.92]	

Appendix 2—table 2. Model parameters for viral dynamics in both the nasopharynx and the trachea estimated by the model with the viral infectivity adjusted for ACE2-RBD-binding inhibition and the loss rate of infected cells adjusted for the group of treatment.
For the bootstrap procedure, 50 iterations were performed.

Parameters	Meaning	Value [95% CI]	Unit
β	Infection rate with ECLRBD = 0 AU ($\times 10^{-8}$)	0.82 [0.13; 5.13]	(copies/mL)$^{-1}$ day^{-1}
	Fold $\Delta\text{ECLRBD} = 10^3$ AU	1.017 [1.012; 1.022]	

Appendix 2—table 2 Continued on next page

Appendix 2—table 2 Continued

Parameters	Meaning	Value [95% CI]	Unit
δ	Loss rate of infected cells	1.02 [0.80; 1.30]	day^{-1}
	Fold change in the convalescent group	1.74 [1.24; 2.46]	
	Fold change in the Conv-CD40 group	2.17 [0.82; 5.74]	
P^N	Viral production rate in the naso. ($\times 10^3$)	8.92 [0.42; 191]	virions $(cell\ day)^{-1}$
P^T	Viral production rate in the trachea ($\times 10^3$)	0.62 [0.02; 19.7]	virions $(cell\ day)^{-1}$
α_{vlsg}	Infected cells and sgRNA viral load ratio	1.32 [0.91; 1.90]	virions $cell^{-1}$
k	Eclipse rate	3	day^{-1}
c	Clearance of de novo produced viruses	3	day^{-1}
c_I	Clearance of inoculum	20	day^{-1}
μ	Percentage of infectious viruses	10^{-3}	
$T_0^{X,nbc}$	Initial number of target cells	1.25×10^5 (naso.) 2.25×10^4 (trachea)	cells
$Inoc_0$	Number of virions inoculated	2.19×10^{10}	virions
ω_β	SD of random effect on $\log_{10} \beta$	0.205 [0.011; 0.399]	
ω_δ	SD of random effect on δ	0.079 [-0.092; 0.250]	
σ_{VLn}	SD of error model gRNA in naso.	1.13 [0.90; 1.36]	
σ_{VLt}	SD of error model gRNA in trachea	1.27 [1.07; 1.48]	
σ_{sgVLn}	SD of error model sgRNA in naso	1.62 [1.30; 1.94]	
σ_{sgVLt}	SD of error model sgRNA in trachea	1.36 [1.15; 1.56]	

6.7 Prague et al. 2016 (Biometrics) Doubly Robust Estimators in Cluster Randomized Trials

Accounting for interference variables using semi-parametric augmentation for improving efficiency in clustered randomized trials with missing at random outcomes. Prague M., Wang R., Stephens A., Tchetgen Tchetgen E and DeGruttola V. *Biometrics.* 72(4) - 1066-1077 - April 2016.

This article integrates in the third axis of my research "Outreach to Implementation : Evaluation in Population". It has been written during my Postdoctoral fellowship at Harvard School of Public Health.

I chose this paper for its significant contribution to causal inference. It introduces a doubly robust GEE estimator that effectively addresses missing data and imbalances in baseline covariates in CRT. The estimator's properties were rigorously tested through simulations and applied to an HIV prevention study in South Africa. Additionally, an R package has been released for public use. The primary challenge in this project was demonstrating the properties of this estimator (bias and variance), a critical aspect for its reliability and application in practical research.

Biometrics 72, 1066–1077
December 2016

DOI: 10.1111/biom.12519

Accounting for Interactions and Complex Inter-Subject Dependency in Estimating Treatment Effect in Cluster-Randomized Trials with Missing Outcomes

Melanie Prague,[1,*] Rui Wang,[1,2] Alisa Stephens,[3] Eric Tchetgen Tchetgen,[4] and Victor DeGruttola[1]

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, U.S.A.
[2]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, Massachusetts, U.S.A.
[3]Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.
[4]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, U.S.A.
email: mprague@hsph.harvard.edu

Summary. Semi-parametric methods are often used for the estimation of intervention effects on correlated outcomes in cluster-randomized trials (CRTs). When outcomes are missing at random (MAR), Inverse Probability Weighted (IPW) methods incorporating baseline covariates can be used to deal with informative missingness. Also, augmented generalized estimating equations (AUG) correct for imbalance in baseline covariates but need to be extended for MAR outcomes. However, in the presence of interactions between treatment and baseline covariates, neither method alone produces consistent estimates for the marginal treatment effect if the model for interaction is not correctly specified. We propose an AUG–IPW estimator that weights by the inverse of the probability of being a complete case and allows different outcome models in each intervention arm. This estimator is doubly robust (DR); it gives correct estimates whether the missing data process or the outcome model is correctly specified. We consider the problem of covariate interference which arises when the outcome of an individual may depend on covariates of other individuals. When interfering covariates are not modeled, the DR property prevents bias as long as covariate interference is not present simultaneously for the outcome and the missingness. An R package is developed implementing the proposed method. An extensive simulation study and an application to a CRT of HIV risk reduction-intervention in South Africa illustrate the method.

Key words: Augmentation; Cluster-randomized trials; Generalized estimating equation (GEE); Interactions; Interference; Inverse probability weighting (IPW); Missing at random (MAR); Outcome Model; Propensity Score; R package; Semi-parametric methods.

1. Introduction

In clustered randomized clinical trials (CRTs), the unit of treatment assignment is a cluster of subjects, which we also refer to as a community. In such settings, outcomes are likely to be correlated among subjects within the same cluster. Often used for estimation, generalized estimating equations (GEE) based on semi-parametric methods (Zeger and Liang, 1986) target marginal effects of treatment. Within clusters, dependence can be modeled using a working correlation structure. Compared to mixed effects models, this approach has the advantage of focusing on population average effects rather than cluster specific effects (which are equal for continuous outcomes) and requires fewer parametric assumptions on the outcome distribution (Hubbard et al., 2010). Moreover, because both the outcome and the missing data mechanism can be modeled, this approach allows doubly robust estimation, which is impossible with mixed effect models. Finally, this approach to estimation is robust to misspecification of the correlation structure. However, challenges arise in developing a consistent and efficient estimator of marginal treatment effects; these include the need to adjust for missing data and accommodate covariate interference (wherein a subject's outcome may be affected by covariates of other subjects) and interactions (wherein the effect of treatment varies by covariate-defined subgroups). We propose a method that addresses these issues and is practical to implement for evaluating novel interventions in CRTs.

In CRTs, covariates may be fully observed even if the outcome is missing. When data are assumed missing completely at random (MCAR)–i.e., the observed process is independent of observed and unobserved information (Rubin, 1976)—the standard GEE approach provides consistent and asymptotically normal (CAN) estimators. If the pattern of missingness depends on observed information but not on missing data, the data are said to be Missing at Random (MAR). In this setting, the standard GEE may yield biased estimates although likelihood-based approaches, such as mixed effect models, can provide unbiased estimators. Imputation (Paik, 1997) or reweighing (Robins et al., 1995) methods can correct for this bias. Although useful if the missingness mechanism is not

completely known, multiple imputation requires correct specification of the joint distribution of the outcomes, which is especially difficult when they are correlated and the cluster sizes are large (Beunckens et al., 2008). In this article, we consider the Inverse Probability Weighting (IPW) approach to analyze incomplete data. If the model for the missingness mechanism represents the MAR data-generating process, the IPW estimation provides CAN estimators of treatment effects by reweighing complete cases according to the probability of being observed (Liang and Zeger, 1986; Robins et al., 1994).

Recent methodological developments improve estimation efficiency by leveraging baseline covariates; they may be based on targeted maximum likelihood (Moore and van der Laan, 2009) and on augmentation (Robins et al., 1994; Robins, 2000; Tsiatis et al., 2008; Zhang et al., 2008). Stephens et al. (2012) developed the augmented GEE (AUG) methods in the setting of dependent outcomes such as in CRTs. The AUG adds a term to the standard GEE which relates the outcome to covariates and treatment. Without missing data, randomization assures that the AUG is CAN. However, in the case of outcome data that are MAR but not MCAR, the AUG may be biased. There exists theory for extending these methods to MAR data for individual randomized Trials (RTs) with possibly correlated data (Van der Laan and Robins, 2003; Glynn and Quinn, 2010); we focus on the details of implementing the methods in CRTs.

The term interference can refer to different types of relationships among exposures, outcomes, and covariates. Interference in RTs arises when one subject's treatment may impact the outcomes of other subjects (Rosenbaum, 2007; Vansteelandt, 2007; Hudgens and Halloran, 2012; Tchetgen Tchetgen and VanderWeele, 2012). A similar phenomenon, confounding by clusters, has been discussed in the context of observational studies (Seaman et al., 2014); we will refer to such confounding as exposure interference. In CRTs, all subjects within a cluster receive the same treatment; hence, if the clusters are independent as typically assumed in practice, there is no exposure interference measured at the cluster level. Therefore, any choice of working correlation structure for the standard GEE will give a consistent estimator of the marginal treatment effect (Pepe and Anderson, 1994). We will investigate covariate interference among individuals nested within clusters: the setting in which one subject's covariate may impact the outcomes of other subjects.

The IPW and the AUG can be combined in a doubly robust method we refer to as the DR; we investigate its properties regarding robustness to misspecification of the missing data and outcome-generating process. By considering a variety of data-generating mechanisms, we investigate settings in which the DR has advantageous properties (consistency and precision) compared to the IPW and the AUG, and discuss the impact of covariate interference and treatment–covariate interactions. This article is organized as follows. Section 2 introduces notation and assumptions for the IPW and the AUG GEE approaches. Section 3 describes the DR approach, investigates CAN properties, and discusses the issue of covariate interference. Section 4 provides a motivating example with data arising from a CRT of an HIV/Sexually Transmitted Infection (STI) risk reduction intervention in South Africa (Jemmott III et al., 2014). Simulation studies regarding bias, relative efficiency, and coverage are described in Section 5, and concluding remarks are made in Section 6.

2. Notation, Basic Models, and Assumptions

2.1. *Notation for CRTs and Marginal Treatment Effect*

We consider a study design in which a vector of P baseline covariates $\boldsymbol{X}_{ij} = (X_{ij}^1, \ldots, X_{ij}^P)$ and outcome Y_{ij} are recorded for each subject $j = 1, \ldots, n_i$ in community $i = 1, \ldots, M$. The sample size within each community is assumed fixed by design and noninformative. Our setting compares two arms (treated $A_i = 1$ and control $A_i = 0$); the probability of treatment assignment is known and given by $p = P(A_i = 1)$; extension to a greater number of treatments is straightforward but complicates the notation. In this article, the outcome $\boldsymbol{Y}_i = [Y_{ij}]_{j=1,\ldots,n_i}$ is assumed to be continuous, but extension to other types of outcomes is straightforward. The vector $\boldsymbol{R}_i = [R_{ij}]_{j=1,\ldots,n_i}$ is the indicator of missingness; Y_{ij} is observed when $R_{ij} = 1$. The matrix of covariates $\boldsymbol{X}_i = [\boldsymbol{X}_{ij}]_{j=1,\ldots,n_i}$ is assumed to be fully observed and consists only of pre-exposure covariates measured at baseline.

Interest lies in estimating the marginal effect of the treatment given by $M_E^* = E(E(Y_{ij}|A_i = 1, \boldsymbol{X}_i) - E(Y_{ij}|A_i = 0, \boldsymbol{X}_i))$. For estimating M_E^*, we make inference about the parameters $\boldsymbol{\beta} = (\beta_0, \beta_A)^T$ indexing the marginal model $g(\mu_{ij}(\boldsymbol{\beta}, A_i)) = g(E(Y_{ij}|A_i)) = \beta_0 + \beta_A A_i$, where $\boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i) = [\mu_{ij}(\boldsymbol{\beta}, A_i)]_{j=1,\ldots,n_i}$ and g is a one-to-one link function, which is an identity function in this article. Of particular interest, β_A is equal to M_E^*. Of note, extension to binary outcome Y_{ij} using a logistic function for g and considering odd-ratios is based on the same reasoning.

When the outcome is believed to be MCAR, the missingness process is independent of $\boldsymbol{X}_i$, A_i, and $\boldsymbol{Y}_i$. If one assumes MAR and the missingness pattern is monotone, the probability of missingness can be estimated by a multistep approach by decomposing a monotone missing pattern into multiple uniform missing data models (Robins et al., 1994; Li et al., 2013). In CRTs, any component of $\boldsymbol{Y}_i$ can be missing; hence, the missingness pattern is nonmonotone. Therefore, we make a stronger assumption than MAR that we refer to as restricted MAR (rMAR): the probability that the outcome for one individual is missing is independent of all outcomes in the cluster, conditional on baseline exposure A_i, and cluster characteristics $\boldsymbol{X}_i$. The conditional probability that the outcome is observed is denoted $\pi_{ij}(\boldsymbol{X}_i, A_i) = P(R_{ij} = 1|\boldsymbol{X}_i, A_i)$ and is called the propensity score (PS). When data are rMAR, ignoring missing data leads to biased inference if missingness depends both on $\boldsymbol{X}_i$ and A_i. This is because the presence of missing data no longer assures balance of confounding factors between treatment arms. Therefore, analysis must include adjustment for missing data; appropriate models for this adjustment may require treatment–covariate interactions, which may be difficult to specify and require many parameters. Combining the IPW and the AUG, which this article proposes, makes it possible to obtain consistent estimates of the marginal effect of treatment without explicitly specifying interaction terms while also improving efficiency.

2.2. *Inverse Probability Weighted Generalized Estimating Equations (IPW)*

In order to account for missing data, semi-parametric estimators based on the IPW are found by solving the estimating equation (1):

$$0 = \sum_{i=1}^{M} \underbrace{\boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W) \left[\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right]}_{\boldsymbol{\psi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W)}, \quad (1)$$

where $\boldsymbol{D}_i = \frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)}{\partial \boldsymbol{\beta}^T}$ is the design matrix, $\boldsymbol{V}_i$ is the covariance matrix equal to $\boldsymbol{U}_i^{1/2} \boldsymbol{C}(\boldsymbol{\alpha}) \boldsymbol{U}_i^{1/2}$ with $\boldsymbol{U}_i$ a diagonal matrix with elements $\text{var}(y_{ij})$, and $\boldsymbol{C}(\boldsymbol{\alpha})$ is the working correlation structure with nondiagonal terms $\boldsymbol{\alpha}$. For example, for an independence correlation structure $\boldsymbol{\alpha}$ are zero; for exchangeable all the elements of $\boldsymbol{\alpha}$ are identical. Parameters $\boldsymbol{\alpha}$ could also depend on the treatment group $\boldsymbol{C}(\boldsymbol{\alpha}(A_i))$ but we do not consider this possibility in our implementation. In this article, we estimate the $\boldsymbol{\alpha}$ parameters using moment estimators from the Pearson residuals as in McDaniel et al. (2013). The $n_i \times n_i$ matrix of weights is $\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W) = \text{diag}\left[R_{ij}/\pi_{ij}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)\right]_{j=1,\ldots,n_i}$, where the PS is obtained by fitting a binary response model that regresses the indicator R_{ij} on functions of A_i and $\boldsymbol{X}_{ij}$. The $\boldsymbol{\eta}_W$ are nuisance parameters estimated in the PS. A necessary assumption for this method is that probabilities for the PS are bounded away from zero. Several authors have noted the instability that may arise from small probabilities of observation (i.e., large weights) and proposed use of stabilized or truncated weights; see Seaman and White (2013) for a review. To ensure that the IPW provides a CAN estimator, the PS must include all variables that are associated simultaneously with both the missingness and outcome processes (Brookhart et al., 2006) including treatment–covariate interaction terms (Belitser et al., 2011). In other words, the PS must be correctly specified, in the sense that $\pi_{ij}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W) = P(R_{ij} = 1 | \boldsymbol{X}_i, A_i)$ for some $\boldsymbol{\eta}_W$.

2.3. *Augmented Generalized Estimating Equations (AUG)*

For settings with complete data, Stephens et al. (2012) proposed the AUG estimator which can improve efficiency relative to the standard GEE by incorporating baseline covariates. The AUG is constructed by subtracting from the set of GEEs the orthogonal projection of the standard estimating function onto the span of scores corresponding to all smooth parametric models for the treatment assignment mechanism given covariates. The AUG is given in equation (2):

$$\begin{aligned} 0 = \sum_{i=1}^{M} \Bigg[& \underbrace{\boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i)\right)}_{\tilde{\boldsymbol{\psi}}_i(\boldsymbol{Y}_i, A_i, \boldsymbol{\beta})} \\ & + \sum_{a=0,1} p^a (1-p)^{1-a} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) \\ & - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a) \Big) \Bigg]. \end{aligned} \quad (2)$$

The term $\tilde{\boldsymbol{\psi}}_i(\boldsymbol{Y}_i, A_i, \boldsymbol{\beta})$ is similar to $\boldsymbol{\psi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W)$ in equation (1) for the IPW except that $\boldsymbol{W}_i$ is set to identity because there is no adjustment for missing data. Definitions for $\boldsymbol{D}_i$ and $\boldsymbol{V}_i$ remain the same. The vector $\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) = \left[B_{ij}(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B)\right]_{j=1,\ldots,n_i}$ is an arbitrary function of $\boldsymbol{X}_i$ given for each treatment arm. The $\boldsymbol{\eta}_B$ are nuisance parameters that must be estimated. The estimator in equation (2) is most efficient if $B_{ij}(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B)$ is equal to $E(Y_{ij} | \boldsymbol{X}_i, A_i = a)$ (Robins et al., 1994; Zhang et al., 2008). For this reason, we shall refer to $\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B)$ as the outcome model (OM), and describe the OM as correctly specified when $B_{ij}(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) = E(Y_{ij} | \boldsymbol{X}_i, A_i = a)$ for some $\boldsymbol{\eta}_B$. In the absence of missing data, the AUG remains consistent even if the OM is not correctly specified. Correct specification can lead to substantial efficiency gains compared to the standard GEE. Moreover, in presence of treatment–covariate interactions, it is useful to fit a different regression model for the OM for each treatment group, e.g., $B_{ij}(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) = \gamma_0^a + \sum_{r=1}^{P} \gamma_r^a X_{ij}^r$ with $\boldsymbol{\eta}_B = (\gamma_1^0, \ldots, \gamma_P^0, \gamma_1^1, \ldots, \gamma_P^1)$, thereby obviating the need to fit covariate–treatment interactions terms. In presence of rMAR, the AUG does not ensure consistent estimation; instead, one must combine the AUG with the IPW as we show below.

3. Methods to Accommodate Missing Data, Treatment Covariate Interactions, and Covariate Interference in CRTs

3.1. *Doubly Robust Augmented IPW Generalized Estimating Equations (DR)*

We extend the AUG in equation (2) to account for missing data using the IPW in equation (1) by subtracting from the set of GEEs the orthogonal projection of $\boldsymbol{\psi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W)$ onto the span of scores corresponding to all smooth parametric models for the missing data process and the treatment assignment mechanism given covariates (Tsiatis, 2006). This gives the following estimating equation (see Web-Supplementary Material B for details):

$$\begin{aligned} 0 = \sum_{i=1}^{M} \Bigg[& \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W) \left(\boldsymbol{Y}_i - \boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B)\right) \\ & + \sum_{a=0,1} p^a (1-p)^{1-a} \boldsymbol{D}_i^T \boldsymbol{V}_i^{-1} \Big(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) \\ & - \boldsymbol{\mu}_i(\boldsymbol{\beta}, A_i = a) \Big) \Bigg], \end{aligned} \quad (3)$$

$$= \sum_{i=1}^{M} \boldsymbol{\Phi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W, \boldsymbol{\eta}_B).$$

The $\boldsymbol{D}_i$, $\boldsymbol{V}_i$ and the PS are defined such as in equation (1), the OM denoted $\boldsymbol{B}_i(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B)$ is defined for each treatment group such as in equation (2). The estimator denoted $\hat{\beta}_{\text{aug}}$ is found by solving the estimating equation given in equation (3). Although analytic solutions sometimes exist, coefficient estimates are generally obtained using an iterative procedure such as the Newton–Raphson method. To get $\hat{\boldsymbol{\beta}}_{\text{aug}}$, we use the estimated PS ($\pi_{ij}(\boldsymbol{X}_i, A_i, \hat{\boldsymbol{\eta}}_W)$) and estimated OM

$(\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \hat{\boldsymbol{\eta}}_B))$. As mentioned above, treatment–covariate interactions can be accounted for by fitting OM regressions separately by treatment group. One could also estimate parameters of the PS model separately by treatment groups. This approach, however, may provide less stable results due to variability in the calculation of weights. In this article, $\hat{\boldsymbol{\eta}}_W$ in $\pi_{ij}(\boldsymbol{X}_i, A_i, \hat{\boldsymbol{\eta}}_W)$ are obtained using a logistic regression and $\hat{\boldsymbol{\eta}}_B$ in $\boldsymbol{B}_i(\boldsymbol{X}_i, A_i, \hat{\boldsymbol{\eta}}_B)$ are obtained using a linear regression. Thus, we treat R_{ij} and $R_{ij'}$ as conditionally independent given A_i and $\boldsymbol{X}_i$. In the presence of correlation of R_{ij} and $R_{ij'}$, one might be able to improve efficiency of estimation of π_{ij} and therefore of the marginal treatment effect by accounting for this correlation. Of note, estimation procedures other than generalized linear models could also be used to compute the OM and the PS values. The DR estimator is doubly robust in the sense that it is CAN under correct specification of either the OM (i.e., $B_{ij}(\boldsymbol{X}_i, A_i = a, \boldsymbol{\eta}_B) = E(Y_{ij}|A_i = a, \boldsymbol{X}_i)$ for some $\boldsymbol{\eta}_B$) or the PS (i.e., $\pi_{ij}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W) = P(R_{ij} = 1|\boldsymbol{X}_i, A_i)$ for some $\boldsymbol{\eta}_W$) (see Web-Supplementary Material Section C1). Implementation in R is available on the CRAN in the package *"CRTgeeDR."* Source code had been made available as Web-Supplementary material. We note that in contrast with several existing software packages (for example, proc GENMOD in SAS, 2015), our implementation of the weighted GEE, which uses $\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ instead of $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$, guarantees consistency for all choices of working correlation structure (see details in Web-Supplementary Material Section C2 and D).

3.2. *Variance of the DR Estimator*

The variance of $\hat{\boldsymbol{\beta}}_{\text{aug}}$ is estimated by the sandwich variance estimator. There are two external sources of variability that need to be accounted for: estimation of $\boldsymbol{\eta}_W$ for the PS and of $\boldsymbol{\eta}_B$ for the OM. We denote $\boldsymbol{\Omega} = (\boldsymbol{\beta}, \boldsymbol{\eta}_W, \boldsymbol{\eta}_B)$ the estimated parameters of interest and nuisance parameters. We can stack estimating functions and score functions for $\boldsymbol{\Omega}$:

$$\boldsymbol{U}_i(\boldsymbol{\Omega}) = \begin{pmatrix} \boldsymbol{\Phi}_i(\boldsymbol{Y}_i, \boldsymbol{R}_i, A_i, \boldsymbol{X}_i, \boldsymbol{\beta}, \boldsymbol{\eta}_W, \boldsymbol{\eta}_B) \\ \boldsymbol{S}_i^W(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W) \\ \boldsymbol{S}_i^B(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_B) \end{pmatrix},$$

where $\boldsymbol{S}_i^W$ and $\boldsymbol{S}_i^B$ represent the score equations for patients in cluster i for the estimation of $\boldsymbol{\eta}_W$ and $\boldsymbol{\eta}_B$ in the PS and the OM. A standard Taylor expansion paired with Slutzky's theorem and the central limit theorem provide the sandwich estimator adjusted for nuisance parameters estimation in the OM and PS. We refer to this as the nuisance-adjusted sandwich estimator:

$$\text{Var}(\boldsymbol{\Omega}) = E\left[\frac{\partial \boldsymbol{U}_i(\boldsymbol{\Omega})}{\partial \boldsymbol{\Omega}}\right]^{-1^T} \underbrace{E\left[\boldsymbol{U}_i(\boldsymbol{\Omega})\boldsymbol{U}_i^T(\boldsymbol{\Omega})\right]}_{\boldsymbol{\Delta}_{\text{adj}}} \underbrace{E\left[\frac{\partial \boldsymbol{U}_i(\boldsymbol{\Omega})}{\partial \boldsymbol{\Omega}}\right]^{-1}}_{\boldsymbol{\Gamma}_{\text{adj}}^{-1}}. \tag{4}$$

The variance estimator $\widehat{\text{var}}(\hat{\boldsymbol{\beta}}_{\text{aug}})$ is obtained by estimating unknown quantities upon substituting empirical means for expectations and $\widehat{\boldsymbol{\Omega}} = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\eta}_W}, \widehat{\boldsymbol{\eta}_B})$ for $\boldsymbol{\Omega}$. Thus, the term $\widehat{\Delta_{\text{adj}}}$ is given by $\frac{1}{M}\sum_{i=1}^{M} \widehat{\boldsymbol{U}}_i(\widehat{\Omega})\widehat{\boldsymbol{U}}_i(\widehat{\Omega})^T$ and $\widehat{\boldsymbol{\Gamma}_{\text{adj}}}$ is given by $\frac{1}{M}\sum_{i=1}^{M} \frac{\partial \widehat{\boldsymbol{U}}_i(\widehat{\boldsymbol{\Omega}})}{\partial \boldsymbol{\Omega}}$.

In small sample settings, it is likely that this estimator of the variance of $\hat{\boldsymbol{\beta}}_{\text{aug}}$ is biased. We implemented Fay's bias-correction approach, which is particularly suitable for M-estimators (Fay et al., 2001). The term $\widehat{\boldsymbol{\Delta}_{\text{adj}}}$ in equation (4) is replaced by $\widehat{\boldsymbol{\Delta}_{\text{fay}}}$ given by $\frac{1}{M}\sum_{i=1}^{M}\left[\widehat{\boldsymbol{H}_i}\widehat{\boldsymbol{U}}_i(\widehat{\boldsymbol{\Omega}})\left(\widehat{\boldsymbol{H}_i}\widehat{\boldsymbol{U}}_i(\widehat{\boldsymbol{\Omega}})\right)^T\right]$, where $\widehat{\boldsymbol{H}_i}$ is a diagonal matrix with diagonal terms $\widehat{\boldsymbol{H}}_{i[jj]} = \left[1 - \min(q, (\frac{\partial \widehat{\boldsymbol{U}}_i(\widehat{\boldsymbol{\Omega}})}{\partial \boldsymbol{\Omega}}\widehat{\boldsymbol{\Gamma}_{\text{adj}}^i})_{[jj]}\right]$, $q = 0.75$ is a frequently used bound.

3.3. *Definition of Covariate Interference and Implication for Analysis*

In previous sections, we discussed covariates measured on the index subject (j), but other subjects' (j') covariates may also impact the outcome for the index subject. An example of a potentially interfering covariate is described by Kaiser et al. (2011) who found a positive association between age of partner and infection with HIV. Similarly, the characteristics of subgroups to which the index case belongs (household, neighborhoods, etc.), whether known or not, may be interfering covariates (Brumback and He, 2011). In this article, we consider the phenomenon of covariate interference where there exists at least one individual $j' \neq j$ such that $E(Y_{ij}|\boldsymbol{X}_{ij}) \neq E(Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{X}_{ij'})$. That is, even after all covariates for the index subject j have been included in the model, the covariates of individuals other than the index subject still affect the outcome of the index subject j; we refer to such covariates as interfering covariates. See Pepe and Anderson (1994) for a similar definition in longitudinal data and see Liu and Hudgens (2014), Seaman et al. (2014) for an analogous definition in nonrandomized clustered data in the context of confounding by cluster and interference. Refer to Web-Supplementary Material Section A for a causal interpretation of covariate-interference.

When interfering covariates affect either the outcome ($E(Y_{ij}|\boldsymbol{X}_{ij}) \neq E(Y_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{X}_{ij'})$) or the missingness process ($E(R_{ij}|\boldsymbol{X}_{ij}) \neq E(R_{ij}|\boldsymbol{X}_{ij}, \boldsymbol{X}_{ij'})$), but not both, the DR estimator is CAN even if the interfering covariates are not included in the models, provided that either the PS or the OM is correctly specified. Accounting for covariate interference in the OM increases efficiency if and only if interfering covariates predict the outcome. When such covariates impact both the outcome and the missing data-generating processes, they must be included in either the OM or the PS models. Thus, the DR estimator is CAN if the model for either the OM or the PS is correctly specified; i.e., either the PS or the OM includes all the covariates $\boldsymbol{X}_i$ in a model that correctly represents the data-generation processes. We acknowledge that this model for interfering covariates is not likely to be known and can be difficult to identify. Different cluster sizes and sub-clustering structures (such as households) may make infeasible the use of regression techniques in the OM or the PS because of the potentially different dimensions of the individual and interfering covariates. Cluster summary measures such as the mean or maximum of individual covariates in the cluster (or

sub-groups in each cluster) may nonetheless be useful in incorporating interference covariates in models (Brumback et al., 2010).

4. Application

4.1. *Description of the SAM Study*

We analyze data from the "South African Men" (SAM) study which randomized 22 pair-matched clusters to a health-promotion intervention (control) and an HIV/STI risk-reduction intervention in a CRT design; the study included 1181 South African men who have sex with women. A complete description of the study design can be found in (Jemmott III et al., 2014). We focus on a cross-sectional analysis of these data after 1 year and ignore matching. The primary outcome of our analysis is the overall percentage of acts of protected intercourse among the total number of acts of intercourse. When the total number of acts of intercourse is zero, we set the percentage to 100%, as no exposure implies no risk. Secondary outcomes are the percentages of protected acts of intercourse by type of partnership and type of intercourse (vaginal and anal sex with main and casual partners). Descriptive statistics for these outcomes, including proportion of missing observations by type of partner and intercourse are provided in Table 1. Slightly more observations are missing in the HIV/STI intervention group (20.8% versus 17.5%). The overall protection percentage after 1 year is about 64% for the HIV/STI intervention compared to 60% for the control group.

As the proportion of missing baseline covariates was less than 0.1%, we consider them to be MCAR and exclude observation with missing covariates from the analysis. No community sub-structure, such as household or neighborhood structures, was described in the SAM study. Here, we consider potential interfering covariates at a cluster level by taking the mean (or mode for qualitative variables) of baseline measures in the community: $\overline{X^k_{i.}} = \frac{1}{n_i}\sum_{j=1,\ldots,n_i} X^k_{ij}$. For example, Hawkes et al. (2013) demonstrated that the mean religiosity score for a community, defined as the mean of individual religiosity score in the community, may have an impact on each individual outcome and missingness in particular regarding sexual behaviors. Table 1 describes socio-demographical individual-level variables and interfering covariates. We provide p-values for Wald tests testing the association of covariates and treatment–covariate interactions with the outcome and the missingness indicator. In this study, there is evidence of interactions of individual covariates with treatment for both the outcome and the missing data-generation processes. However, the interfering covariates defined here do not appear to be significantly associated with both the outcome and the missing data-generation process.

4.2. *Results*

We analyze these data with the GEE, the AUG, the IPW, and the DR using both independence (-I) and exchangeable (-E) working correlation structures. Variables for the PS, and the OM were selected using a forward stepwise regression (separately for each treatment group) from among all the individual covariates $\boldsymbol{X}_{ij}$ presented in Table 1. We did not include the interfering covariates ($\overline{X_{i.}}$) in the analysis as none impacted both outcome and missingness processes (Table 1). We used the *step* function in R based on the AIC criterion. Results of these selections are given in Web-Supplementary Material F. We describe here the results for the primary outcome. The amount of missingness is larger in the treated arm and increases with age; it decreases with religiosity, good health score, and exercise. The OM patterns are substantially different for treated and control; the only common variable is the CAGE score. In both arms, lower alcohol consumption is associated with a greater percentage of protected acts of intercourse. Results are presented in Table 2 for primary and secondary outcomes. With the DR-E, we observe a significant difference of 7.4% (sd = 2.9%, p = 0.01) in the overall percentage of protected intercourse in the HIV/STI intervention group compared to the control group. Analyses of the secondary outcomes suggest that this result is mainly driven by condom use during vaginal intercourse with a marital partner. The HIV/STI intervention has no significant impact on other outcomes. Using the DR rather than the standard GEE or the AUG has an impact on the treatment effect estimates and associated standard errors (SE). The difference between these approaches is apparent in the magnitude and direction of the marginal treatment effect estimate. For example, the analysis for the GEE-I (3.8 [−1.0; 8.5]) does not demonstrate a significant effect of the HIV/STI intervention on overall percentage of protected intercourse, whereas this effect is stronger and significant for the DR-I (7.3 [1.6; 13.0]). Both the GEE-I and the AUG-I (5.4 [2.2; 8.7]) are probably biased due to missing data. Using the DR instead of the IPW leads to an increased magnitude of the treatment effect and an increased level of statistical significance: for example, the DR-E (7.4 [1.73; 13.0]) compared to the IPW-E (3.4 [-1.4; 8.3]).

5. Simulation Studies

5.1. *Properties of the DR Estimator*

We consider a setting with continuous outcome Y_{ij} and assignment of treatment A_i at a cluster level with probability $p = 1/2$. We generate a normally distributed covariate X^1_{ij} (independent of A_i) with mean 1 and a standard deviation of 5. For each individual, we define a covariate $\overline{\boldsymbol{X}^1_{i.}}$ which is the mean of $\boldsymbol{X}^1$ for all the subjects in the same cluster: $\overline{\boldsymbol{X}^1_{i.}} = \frac{1}{n_i}\sum_{j=1}^{n_i} X^1_{ij}$. Similarly, we generate $X^2_{ij} \sim \mathcal{N}(2, 5)$ and $X^3_{ij} \sim \mathcal{N}(3, 5)$; $\overline{\boldsymbol{X}^2_{i.}}$ and $\overline{\boldsymbol{X}^3_{i.}}$ are defined as was $\overline{\boldsymbol{X}^1_{i.}}$ and are possible interfering covariates. The model for simulation is given in equation (5):

$$\begin{cases} Y_{ij} & = \beta^O_0 + \beta^O_A A_i + \beta^O_1 X^1_{ij} \\ & \quad + \beta^O_{I1}\overline{X^1}_{i.} + \beta^O_{A1} A_i X^1_{ij} + \epsilon^O_i + \epsilon^O_{ij} \\ \text{logit}(P(R_{ij}=0)) & = \beta^M_0 + \beta^M_A A_i \\ & \quad + \beta^M_1 X^1_{ij} + \beta^M_{I1}\overline{X^1}_{i.} + \beta^M_{A1} A_i X^1_{ij} \end{cases}. \quad (5)$$

The parameters $\boldsymbol{\beta}^O = (\beta^O_0, \beta^O_A, \beta^O_1, \beta^O_{I1}, \beta^O_{A1})$ are the regressors associated with intercept, treatment, covariate, interfering covariate, treatment–covariate interaction for the outcome

Table 1

Descriptive statistics of outcomes, sociodemographic individual covariates, and interfering covariates by intervention group in SAM study. Values in bold in the table are p-values < 0.1.

Descriptive statistics of the outcomes						
	HIV/STI		Control group			
	Mean [IQR]	% missing	Mean [IQR]	% missing		
Primary outcome for percentage of protection (Y)						
Overall	64% [26; 100]	20.8%	60% [22; 100]	17.5%		
Secondary outcomes for percentage of protection (Y^1, Y^2, Y^3, and Y^4)						
Main partner vaginal sex	61% [22; 100]	10.2%	56% [0; 100]	9.3%		
Casual partners vaginal sex	68% [33; 100]	19.7%	68% [33; 100]	17.1%		
Main partner anal sex	37% [0; 68]	11.2%	52% [0; 100]	8.6%		
Casual partners anal sex	35% [0; 100]	15.1%	31% [0; 100]	12.8%		
Descriptive statistics of the covariates						
			p-value for association with			
	HIV/STI	Control group	Y[a]		P(Y observed)[b]	
	Mean [IQR]	Mean [IQR]	$\eta_2^O \neq 0$	$\eta_3^O \neq 0$	$\eta_2^M \neq 0$	$\eta_3^M \neq 0$
Individual covariates X_{ij}						
Age	26 [21; 30]	26.5 [21; 31]	0.41	0.13	**0.03**	0.18
Employment yes	23%	26%	**0.04**	0.17	**0.01**	**<0.001**
Married yes	23%	24%	**0.05**	0.76	0.68	0.50
Education yes	46%	42%	0.58	**<0.001**	0.76	**0.05**
Number of children	1.5 [0; 2]	1.7 [0; 2]	0.21	0.12	0.25	0.31
Wealth	5.3 [4; 7]	5.3 [4; 7]	0.77	0.96	0.25	0.54
Social desirability	3.4 [3.2; 3.4]	3.4 [3.2; 3.4]	0.87	0.33	**0.04**	0.34
Religiosity	0.01 [−0.7; 0.7]	0.00[-0.7; 0.6]	0.46	0.25	**0.07**	0.69
HIV/STI knowledge	14.3 [12; 17]	14.1 [12; 17]	0.13	0.93	0.37	**0.03**
Condom behaviors	3.7 [3.3; 4]	3.7 [3.3; 4.1]	**<0.001**	0.36	0.16	0.33
Condom knowledge	3.1 [3; 4]	3.1 [3; 4]	0.41	0.57	0.21	**0.06**
Condom efficacy	3.9 [3.7; 4.2]	3.9 [3.7; 4.2]	**0.01**	0.31	0.97	0.42
Condom peer norm	3.7 [3.4; 4.1]	3.7 [3.4; 4]	**<0.001**	0.71	0.49	0.32
Never had HIV test	20%	21%	0.61	0.80	0.74	0.34
Sexual activity Yes	84%	84%	0.71	**0.06**	0.53	0.77
Eating attitude	4.2 [4; 5]	4.2 [3.7; 5]	0.76	**0.01**	0.74	0.53
Exercise yes	43%	42%	0.99	**0.04**	0.12	0.46
CAGE >= 2	62%	58%	0.22	0.41	0.18	**0.08**
Health knowledge	10.8 [9; 12]	10.6 [9; 13]	0.51	0.38	0.59	0.83
Interfering covariates $\overline{X_{i.}} = \frac{1}{n_i}\sum_{j=1,\ldots,n_i} X_{ij}$						
Mean age	26 [25; 27]	27 [26; 28]	0.39	0.96	**0.05**	**0.10**
Mean education yes	27%	8%	0.58	0.61	0.72	1.00
Mean number of children	1.6 [1.2; 2.1]	1.7 [1.1; 2.1]	0.81	0.67	0.14	0.59
Mean wealth	5.4 [4.4; 6.2]	5.2 [4.4; 6.1]	0.45	0.38	0.23	0.92
Mean social desirability	3.4 [3.3; 3.4]	3.4 [3.3; 3.4]	0.16	0.44	0.60	0.85
Mean religiosity	0.00 [−0.1; 0.1]	0.00 [-0.1; 0.1]	0.84	0.70	0.18	0.94
Mean HIV/STD knowledge	14.2 [14; 15]	13.9 [13; 14]	0.37	0.23	**0.01**	0.45
Mean condom behaviors	3.7 [3.6; 3.8]	3.7 [3.7; 3.8]	0.37	0.40	**0.02**	0.95
Mean condom knowledge	3.1 [2.9; 3.3]	3.1 [2.9; 3.2]	0.52	0.21	0.15	0.32
Mean condom efficacy	3.9 [3.7; 4.0]	3.9 [3.8; 4.0]	0.23	0.38	0.21	0.58
Mean condom peer norm	3.7 [3.6; 3.8]	3.7 [3.6; 3.7]	0.23	0.52	**<0.001**	**0.01**
Mean eating attitude	4.2 [4.1;4.3]	4.2 [4.0; 4.3]	0.71	0.15	0.25	**0.07**
Mean exercise yes	76%	82%	0.43	0.53	**0.10**	0.82
Mean CAGE>=2	63%	37%	0.99	0.79	0.71	0.41
Mean health knowledge	10.7 [10.5; 11]	10.6 [10.3; 10.8]	**0.10**	**0.10**	0.15	0.73

[a] Wald test for η_2^O and η_3^O in the regression $Y = \eta_0^O + \eta_1^O A + \eta_2^O X + \eta_3^O AX$

[b] Wald test for η_2^M and η_3^M in the regression $\text{logit}[P(R=1)] = \eta_0^M + \eta_1^M A + \eta_2^M X + \eta_3^M AX$

Table 2
Analysis of effect of STI/HIV intervention on overall percentage of protected intercourses during the last 3 months 1 year after intervention (primary outcome) and stratified by intercourse types (secondary outcomes) in SAM study with the GEE, the IPW, the AUG and the DR

	Independence (-I)			Exchangeable (-E)		
	$\hat{\beta}_A$	SE	p-value	$\hat{\beta}_A$	SE	p-value
			Overall percentage of protected intercourse (Y)			
GEE	3.751	2.419	0.121	3.738	2.361	0.113
IPW	3.445	2.558	0.178	3.429	2.488	0.168
AUG	5.414	1.665	0.001	5.478	1.633	0.001
DR	7.341	2.923	0.012	7.386	2.885	0.010
			Percentage of protected vaginal intercourse with marital partner (Y^1)			
GEE	5.805	2.689	0.031	5.761	2.67	0.031
IPW	5.660	2.720	0.037	5.626	2.698	0.037
AUG	6.550	1.811	<0.001	6.518	1.794	<0.001
DR	7.254	2.542	0.004	7.273	2.50	0.004
			Percentage of protected vaginal intercourse with casual partner (Y^2)			
GEE	−0.621	4.180	0.882	−0.497	4.164	0.905
IPW	−1.500	4.182	0.720	−1.356	4.17	0.745
AUG	−1.191	2.638	0.652	−1.121	2.624	0.669
DR	−2.103	4.077	0.606	−2.018	4.058	0.619
			Percentage of protected anal intercourse with marital partner (Y^3)			
GEE	−0.983	1.083	0.364	−0.972	1.081	0.369
IPW	−0.934	1.087	0.390	−0.921	1.085	0.396
AUG	−0.951	0.684	0.164	−0.954	0.684	0.163
DR	−0.835	1.005	0.406	−0.819	1.003	0.414
			Percentage of protected anal intercourse with casual partner (Y^4)			
GEE	0.013	1.201	0.991	−0.002	1.204	0.998
IPW	−0.003	1.181	0.998	−0.019	1.184	0.987
AUG	−0.467	0.834	0.576	−0.476	0.837	0.570
DR	−0.963	1.207	0.425	−0.971	1.208	0.421

model. Parameters $\boldsymbol{\beta}^M$ are the same for the missing data-generating process. Scenarios with low correlation among cluster (0.05) were simulated with $\epsilon_i^O \sim \mathcal{N}(0, 0.05)$ and $\epsilon_{ij}^O \sim \mathcal{N}(0, 1.0)$ for cluster and individual random errors; scenarios with high correlation (0.2) were simulated with $\epsilon_i^O \sim \mathcal{N}(0, 0.25)$ and $\epsilon_{ij}^O \sim \mathcal{N}(0, 1.0)$. True correlation structure is exchangeable. We investigate small sample ($M = 10$ and $n_i = (10, 20, 30)$ with probability 1/3 each) and large sample ($M = 100$ and $n_i = (90, 100, 110)$ with probability 1/3 each) properties. In each scenario, we generate 1000 replicates of datasets.

We evaluate the double robustness of the DR estimator in the setting of large and small sample with low correlation, but similar results are observed for large correlation. We investigate models of analysis with OM and PS correctly specified (TRUE), misspecified (MISS), and partially specified omitting treatment–covariate interactions (NONE). Table 3 describes the data-generation process, provides the formulations of the models of analysis, and shows the results from analysis; on average, 26% of outcomes were missing and the average ICC was 0.08. When there is no missing data, the traditional GEE is consistent because of randomization. When outcome data are MAR but not MCAR, the GEE and the AUG analysis are biased (−1.7 for the GEE-I and −1.8 for the AUG-I). When either the OM or the PS models or both are correctly specified, there is negligible estimated bias for the DR—a finding that confirms consistency. In small samples, this bias is bigger when only the PS is correct because the weights are estimated with lower accuracy. Using the more common choice of implementation for the weighted GEE $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ leads to very high bias if an exchangeable correlation structure is used (0.374 if the OM is correct and 858 if it is not, for large sample). When the OM is correct, the coverage remains around 95% (see Table 2 in Web-Supplementary Material E). Using $\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ in the implementation of weights addresses this problem and permits the use of correlation structures other than independence. The IPW with correct PS also corrects the bias (−0.01) but is less efficient than the DR approach; coverage

Table 3

Properties for the Doubly robust estimator (DR) compared to the GEE, the IPW, and the AUG using the data-generation mechanism from equation (5) with covariate interference for the outcome and missing data-generation process. Misspecified (.MISS), correctly specified (.TRUE), and partially specified without treatment–covariate interactions (.NONE) OM and PS are investigated. Statistics for 1000 replicates are the bias compared to $M_E^* = 2.0$, the empirical standard errors over the replicates, the mean asymptotic nuisance-adjusted standard error, and the coverage with independence (-I) and exchangeable (-E) working correlation matrix.

				Standard error (SE)				Coverage	
		Bias		Empirical		Robust		95%	
	M_E^*	-I	-E	-I	-E	-I	-E	-I	-E
Small sample $M = 10, n_i = (10, 20, 30)$ with probability 1/3 each, Low correlation									
GEE (no missing)	2.0	0.0186	0.0171	0.6553	0.6598	0.5629	0.5682	93.0	92.9
GEE	2.0	−1.7186	−1.7166	0.5717	0.5724	0.5074	0.4306	12.8	7.2
IPW.PS.TRUE	2.0	−0.1623	−0.1689	1.1447	1.1473	0.7987	0.8161	83.9	84.7
AUG.OM.TRUE	2.0	−1.8142	−1.8134	0.4530	0.4148	0.8751	0.8699	39.4	38.0
DR.OM.MISS.PS.TRUE	2.0	−0.0127	−0.0366	2.7327	2.6793	1.4029	1.3985	92.0	92.0
DR.OM.TRUE.PS.MISS	2.0	0.0011	0.0001	0.1544	0.1545	0.1287	0.1330	86.0	87.5
DR.OM.TRUE.PS.TRUE	2.0	−0.0017	−0.0022	0.1881	0.1838	0.1413	0.1447	86.9	87.4
DR.OM.TRUE.PS.NONE	2.0	0.0006	−0.0003	0.1612	0.1608	0.1330	0.1368	85.8	87.8
Large sample $M = 100, n_i = (90, 100, 110)$ with probability 1/3 each, Low correlation									
GEE (no missing)	2.0	0.0042	0.0043	0.1156	0.1157	0.1155	0.1156	94.3	94.5
GEE	2.0	−1.7335	−1.7321	0.1015	0.1013	0.0994	0.0994	0.0	0.0
IPW.TRUE	2.0	−0.0113	−0.0108	0.2626	0.2621	0.2507	0.2510	93.5	93.9
AUG.TRUE	2.0	−1.8021	−1.8024	0.0694	0.0664	0.2556	0.2550	0.0	0.0
OM.MISS.PS.TRUE	2.0	−0.0089	−0.0079	0.3127	0.3105	0.3937	0.3940	99.3	99.1
OM.TRUE.PS.MISS	2.0	0.0013	0.0014	0.0259	0.0259	0.0256	0.0257	95.2	95.7
OM.TRUE.PS.TRUE	2.0	0.0013	0.0014	0.0284	0.0284	0.0285	0.0285	95.8	96.0
OM.TRUE.PS.NONE	2.0	0.0014	0.0014	0.0266	0.0266	0.0263	0.0263	95.2	95.1

Marginal model for the GEE:

$\mu_{ij}(\boldsymbol{\beta}, A_i) = \beta_0 + \beta_A A_i$

OM is fitted for each treatment group $A_i = a$:

OM.TRUE $B_{ij}(\boldsymbol{X}_i, A_i = a) = \gamma_0^a + \gamma_1^a X_{ij}^1 + \gamma_2^a \overline{X^1}_{i\cdot}$

OM.MISS $B_{ij}(\boldsymbol{X}_i, A_i = a) = \gamma_0^a + \gamma_1^a X_{ij}^2$

PS is fitted for the whole dataset:

PS.TRUE $\pi_{ij}(\boldsymbol{X}_i, A_i) = \text{expit}\left(\gamma_0^M + \gamma_A^M A_i + \gamma_1^M X_{ij}^1 + \gamma_2^M \overline{X^1}_{i\cdot} + \gamma_3^M A_i X_{ij}^1\right)$

PS.MISS $\pi_{ij}(\boldsymbol{X}_i, A_i) = \text{expit}\left(\gamma_0^M + \gamma_A^M A_i + \gamma_1^M X_{ij}^2\right)$

PS.NONE $\pi_{ij}(\boldsymbol{X}_i, A_i) = \text{expit}\left(\gamma_0^M + \gamma_A^M A_i + \gamma_1^M X_{ij}^1 + \gamma_2^M \overline{X^1}_{i\cdot}\right)$

is close to the nominal value of 95%. In small samples, the empirical SE are underestimated. By contrast, in the large sample setting, using the nuisance-adjusted sandwich estimator for the DR leads to good estimates of the asymptotic SE (0.0263) compared to the empirical SE (0.0266) over 1000 replicates. Moreover, we observe that the coverage using the DR is comparable to that of the GEE with complete data. Finally, we note that when the treatment–covariate interactions are ignored in the PS and only accounted for in the OM by fitting a different regression in each treatment group, the DR approach is also consistent and achieves same precision as when both the PS and the OM are correct (0.0014 and SE = 0.027 for OM.TRUE.PS.NONE and 0.0013 SE = 0.029 for OM.TRUE.PS.TRUE).

Table 4 presents the results of analyses with the GEE, the IPW, the AUG, and the DR that investigate the impact of correlation of the outcome in the data with small and large sample. The average percentage of missing outcomes is 23%; the average ICC is 0.04 for low correlation and 0.21 for high correlation. We analyzed the data using a PS and an OM model that was fit using a stepwise variable selection from among all of the individual and interfering covariates

Table 4

Sample size effect and correlation magnitude effects for data-generation mechanism given in equation (5) with $\boldsymbol{\beta}^O = (1, 1, 1, 1, 1)$ *and* $\boldsymbol{\beta}^M = (-3, 1/2, 1/2, 1/2, 1/2)$. *Statistics for 1000 replicates are the bias compared to* M_E^*, *the empirical standard errors over the replicates, the mean asymptotic nuisance-adjusted standard errors and the coverage for the GEE, the IPW, the AUG, and the DR with independence (-I) and exchangeable (-E) working correlation matrix.*

				Standard error (SE)						Coverage			
		Bias		Empirical		Robust		Fay's		Robust		Fay's	
	M_E^*	-I	-E	-I	-E	-I	-E	-I	-E	-I	-E	-I	-E
Small sample $M = 10, n_i = (10, 20, 30)$ with probability 1/3 each, Low correlation													
GEE	2.0	−1.7473	−1.7479	0.4351	0.4360	0.3963	0.3256	0.4559	0.4603	0.8	2.3	3.9	4.9
IPW	2.0	−1.0130	−1.0130	0.6793	0.6842	0.5538	0.5591	0.6735	0.6766	49.0	49.2	59.8	59.9
AUG	2.0	−1.8099	−1.8111	0.3371	0.3269	0.8362	0.8353	0.8834	0.8817	29.7	29.1	40.1	39.2
DR	2.0	0.0008	0.0006	0.1552	0.1586	0.1127	0.1140	0.1190	0.1201	84.8	83.8	86.0	86.2
Large sample $M = 100, n_i = (90, 100, 110)$ with probability 1/3 each, Low correlation													
GEE	2.0	−1.7335	−1.7321	0.1015	0.1013	0.0985	0.0727	0.0994	0.0994	0.0	0.0	0.0	0.0
IPW	2.0	−0.9955	−0.9952	0.1514	0.1517	0.1559	0.1563	0.1588	0.1592	0.2	0.2	0.2	0.2
AUG	2.0	−1.8019	−1.8022	0.0695	0.0664	0.2556	0.2550	0.2569	0.2563	0.0	0.0	0.0	0.0
DR	2.0	0.0016	0.0017	0.0265	0.0265	0.0262	0.0263	0.0264	0.0264	95.1	95.0	95.1	95.2
Small sample $M = 10, n_i = (10, 20, 30)$ with probability 1/3 each, High correlation													
GEE	2.0	−0.0086	−0.0086	0.5265	0.5314	0.4701	0.4721	0.5651	0.5657	88.5	88.4	92.9	92.7
IPW	2.0	−1.0221	−1.0229	0.7026	0.7083	0.5776	0.5829	0.7015	0.7044	52.4	52.2	62.2	61.5
AUG	2.0	−1.7985	−1.7987	0.5058	0.5084	0.8748	0.8727	0.9243	0.9209	35.8	35.8	45.1	45.5
DR	2.0	0.0098	0.0062	0.4328	0.4407	0.2469	0.2480	0.2607	0.2614	77.4	77.7	79.7	79.6
Large sample $M = 100, n_i = (90, 100, 110)$ with probability 1/3 each, High correlation													
GEE	2.0	−1.7325	−1.7312	0.1145	0.1141	0.1121	0.0753	0.1132	0.1132	0.0	0.0	0.0	0.0
IPW	2.0	−0.9945	−0.9940	0.1618	0.1620	0.1652	0.1656	0.1682	0.1686	0.2	0.2	0.2	0.2
AUG	2.0	−1.8014	−1.8017	0.0787	0.0761	0.2587	0.2581	0.2600	0.2594	0.0	0.0	0.0	0.0
DR	2.0	0.0029	0.0032	0.0609	0.0610	0.0590	0.0590	0.0593	0.0593	94.7	94.6	94.7	94.6

Marginal model for the GEE:

$\mu_{ij}(\boldsymbol{\beta}, A_i) = \beta_0 + \beta_A A_i$

OM in AUG and DR is fitted for each treatment group $A_i = a$ using a stepwise regression:

$\boldsymbol{B}_{ij}(\boldsymbol{X}_i, A_i = a) = \text{stepwise}(X_{ij}^1, X_{ij}^2, X_{ij}^3, \overline{X^1}_{i.}, \overline{X^2}_{i.}, \overline{X^3}_{i.})$

PS in DR and IPW is fitted for the whole dataset using a stepwise regression:

$\text{logit}(\pi_{ij}(\boldsymbol{X}_i, A_i)) = \text{stepwise}(\boldsymbol{A}_i, X_{ij}^1, X_{ij}^2, X_{ij}^3, \overline{X^1}_{i.}, \overline{X^2}_{i.}, \overline{X^3}_{i.})$

described above. The GEE and the AUG estimates are systematically biased because there is no correction for missing data. The IPW is also biased because the PS is incorrect in that it omits treatment–covariate interactions. The DR estimates are consistent in all analyses. In small sample settings, the empirical SE is underestimated even when using nuisance-adjusted SE, but estimation is improved by Fay's correction. Nonetheless, the coverage remained lower than 86%, but it improves for large samples. Finally, when there is low correlation in the outcome, the robust SE better approximates the empirical SE.

5.2. *Simulations Mimicking the SAM Study*

To consider more complex settings, we mimic the SAM study (see Section 4). We simulate the following individual-level covariates: employment (EMP $\sim \mathcal{B}(0.25)$), marital status (MARI $\sim \mathcal{B}(0.23)$), age (AGE $\sim \mathcal{N}(27; 7)$), religiosity (REL $\sim \mathcal{N}(0, 0.8)$), the CAGE score (from a multinomial of probabilities CAGE $\sim \mathcal{M}(0.3; 0.1; 0.1; 0.2; 0.3)$ for modalities 0,1,2,3 and 4), the HIV score (HIV $\sim \mathcal{N}(14; 4)$), and the condom knowledge score (CDM $\sim \mathcal{N}(3; 1)$). Interfering covariates are generated as means for quantitative variables or modes for qualitative variables of the individual-level variables in each of the community (as was done for $\overline{\boldsymbol{X}^1}$, $\overline{\boldsymbol{X}^2}$, and $\overline{\boldsymbol{X}^3}$ in Section 5.1). We generate data from the model in equation (6). In simulating the outcome, we add cluster random errors to create an exchangeable correlation structure with $\epsilon_i^O \sim \mathcal{N}(0, 5)$ and an individual random effects $\epsilon_{ij}^O \sim \mathcal{N}(0, 4)$. This provides an outcome correlation among clusters of 0.07. We analyzed the data using a PS and an OM composed of all the covariates

Table 5

Simulation of the scenario described in equation (6) mimicking the SAM study data. Statistics for 1000 replicates are the bias compared to M_E^*, the empirical standard errors over replicates, the mean asymptotic nuisance-adjusted standard error, and the coverage for the GEE, the IPW, the AUG, and the DR with independence (-I) and exchangeable (-E) working correlation matrix.

				Standard error (SE)						Coverage			
		Bias		Empirical		Robust		Fay's		Robust		Fay's	
	M_E^*	-I	-E	-I	-E	-I	-E	-I	-E	-I	-E	-I	-E
Small sample $M = 10, n_i = (10, 20, 30)$ with probability 1/3 each													
GEE	5.73	2.214	2.213	1.330	1.329	1.829	1.848	1.359	1.363	89.4	88.7	59.4	59.7
IPW	5.73	0.536	0.537	1.333	1.333	1.214	1.214	1.470	1.471	86.5	86.4	92.1	92.1
AUG	5.73	0.173	0.173	0.973	0.973	0.878	0.878	0.925	0.925	88.6	88.6	89.9	89.9
DR	5.73	−0.104	−0.104	1.102	1.101	0.932	0.931	0.982	0.982	90.3	90.3	92.0	92.0
SAM-like sample $M = 50, n_i = (20, 30, 30)$ with probability 1/3 each													
GEE	5.73	2.347	2.343	0.308	0.308	0.532	0.466	0.308	0.309	0.0	0.0	0.0	0.0
IPW	5.73	0.622	0.623	0.303	0.303	0.317	0.317	0.323	0.323	50.7	50.7	52.1	52.1
AUG	5.73	0.215	0.215	0.222	0.222	0.230	0.230	0.232	0.232	85.1	85.1	85.2	85.2
DR	5.73	0.037	0.026	0.259	0.260	0.252	0.253	0.254	0.255	94.6	95.3	94.6	95.4

Marginal model for the GEE:

$\mu_{ij}(\boldsymbol{\beta}, A_i) = \beta_0 + \beta_A A_i$

OM in AUG and DR is fitted for each treatment group using a stepwise regression:

$\boldsymbol{B}_{ij}(\boldsymbol{X}_i, A_i = a) = \text{stepwise}(\text{EMP}_{ij}, \text{MARI}_{ij}, \text{AGE}_{ij}, \text{REL}_{ij}, \text{CAGE}_{ij}, \text{HIV}_{ij}, \text{CDM}_{ij}, \text{X}^1{}_{ij}, \text{X}^2{}_{ij}, \text{X}^3{}_{ij})$

PS in IPW and DR is fitted for the whole dataset using a stepwise regression:

$\text{logit}(\pi_{ij}(\boldsymbol{X}_i, A_i)) = \text{stepwise}(\text{A}_i, \text{EMP}_{ij}, \text{MARI}_{ij}, \text{AGE}_{ij}, \text{REL}_{ij}, \text{CAGE}_{ij}, \text{HIV}_{ij}, \text{CDM}_{ij}, \text{X}^1{}_{ij}, \text{X}^2{}_{ij}, \text{X}^3{}_{ij})$

described above with a stepwise variable selection. Table 5 shows the bias, SE, and coverage of the methods we consider based on 1000 replicates for the estimation of the parameter $M_E^* = 5.73$. The percentage of missing outcomes is 21% and the average empirical ICC is 0.06.

$$\begin{cases} Y_{ij} = 60 + 40A_i - 9.0\text{EMP}_{ij} - 8.0\text{MARI}_{ij} + 1.0\text{CDM}_{ij} + 5.0\text{REL}_{ij} \\ \quad + \underbrace{A_i\left[-2.0\text{AGE}_{ij} + 8.5\text{EMP}_{ij} + 3.5\text{MARI}_{ij} + 1.5\text{HIV}_{ij} - 2.0\text{CAGE}_{ij} + 2.0\text{REL}_{ij}\right]}_{\text{Interactions}} \\ \quad \underbrace{-0.5\overline{\text{AGE}_{i\cdot}} - 7.0\overline{\text{CDM}_{i\cdot}} - 5\overline{\text{REL}_{i\cdot}} + 1.0\overline{\text{HIV}_{i\cdot}}}_{\text{covariate interference}} + \epsilon_i^O + \epsilon_{ij}^O \\ \text{logit}[P(R_{ij} = 0)] = -3.0 + 2.0A_i + 0.01\text{AGE}_{ij} - 0.1\text{HIV}_{ij} + \underbrace{A_i\left[-0.1\text{AGE}_{ij} - 0.2\text{HIV}_{ij}\right]}_{\text{Interactions}} \\ \quad + \underbrace{0.02\overline{\text{AGE}_{i\cdot}} + 0.2\overline{\text{CDM}_{i\cdot}} + 0.2\overline{\text{CAGE}_{i\cdot}}}_{\text{covariate interference}} \end{cases} \quad (6)$$

Table 5 provides the estimates the marginal treatment effect for small sample and for the same sample size as that of the SAM data. The GEE, the AUG, and the IPW yield biased results whereas the DR has small bias justifying its use to analyze the data even ignoring covariate interference. Fay's correction with coverage around 92% in small sample and 95% in large sample achieve good accuracy. Figure 2 in Web-Supplementary Material C3 represents the histograms of estimates over the 1000 replicates together with the true value of marginal treatment effect. It displays the bias of the GEE, the AUG, and the IPW estimators compared to the DR and supports the approximate normal distribution of the DR estimator.

6. Discussion

We propose a doubly robust method for the estimation of the marginal effect of treatment in CRTs with continuous data subject to rMAR—an assumption that arises because missingness is nonmonotone in CRTs. Extension to binary or other outcomes is straightforward, provided that there is a one-to-one link function h such that: $\mu_{ij} = h(\boldsymbol{X}_i, A_i)$. We extend the IPW approach proposed by Robins et al. (1995)

and the AUG approach for CRTs proposed by Stephens et al. (2012). To be CAN, the DR estimator requires that either the OM or PS model be correctly specified regardless of the choice of the working correlation matrix. Interfering covariates can be ignored if either the OM or the PS is correctly specified. In presence of treatment–covariate interactions, if the PS is not correctly specified, covariates that interact with treatment on the outcome must be included in the OM. We accommodate these treatment–covariate interactions by modeling the OM separately for each treatment group. Covariates for the OM and the PS may be selected using automatic variable selection procedures such as a stepwise procedure, and may be at the cluster level or individual level.

We recommend using $\boldsymbol{V}_i^{-1}\boldsymbol{W}_i(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ to ensure consistency of the IPW and the DR for CRTs, rather than the conventional implementation, $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ available in several software packages of the weighted GEE. See Tchetgen Tchetgen et al. (2012) for a similar result for longitudinal data with observation-specific weights. If a working independence correlation structure is used, then the two implementations lead to the same result. When $\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)\boldsymbol{V}_i^{-1}\boldsymbol{W}_i^{1/2}(\boldsymbol{X}_i, A_i, \boldsymbol{\eta}_W)$ and an arbitrary correlation structure are used in the DR, estimation of marginal treatment effect is consistent only if the OM is correctly specified. We provide an R package called *CRTgeeDR* that implements the proposed DR estimator. The application of our methods to data from the SAM study showed an effect of HIV/STI intervention on the percentage of protected intercourse (Jemmott III et al., 2014) that reached a 0.05 level of significance. Moreover, results of the analysis that distinguishes among different types of partners and of sexual behavior may be useful in targeting future interventions. Our approach allows a situation that we denoted covariate interference in CRTs, and thus extends ideas of adjustment of time-varying covariates in longitudinal responses (Pepe and Anderson, 1994; Tchetgen Tchetgen et al., 2012). Since treatment is randomized at a cluster level and we consider a marginal mean model which only includes treatment, the covariate interference has a different implication for analysis than exposure interference in causal framework (Liu and Hudgens, 2014) or confounding by cluster in observational studies (Berlin et al., 1999; Huang and Leroux, 2011). However, when there are interactions between X_{ij}^r and A_i exposure and covariate interference are related; in this case, individual ij may be seen as receiving pseudo-treatment $A_i X_{ij}^r$. For such a setting, our work may be seen as extending the notion of exposure interference in RTs to CRTs and is related to the work of Ogburn and VanderWeele (2014). In any case, modeling covariate interference may lead to substantial gains of efficiency if they predict the outcome. Therefore, it may be profitable to develop methods that make use of contact network information to inform the selection of interfering covariates. Finally, the impact of violation of the rMAR assumption required for the consistency of the DR estimates that resulted from an MNAR missingness mechanism can be investigated by performing sensitivity analysis (Rotnitzky et al., 1998; Vansteelandt et al., 2007).

7. Supplementary Materials

Web Appendices, Tables, Figures, simulated data, and R sources implementing the estimators referenced in Sections 3.1, 3.3, 4.2, 5.1 and 5.2 are available with this article at the *Biometrics* website on Wiley Online Library.

Acknowledgements

We thank the associate editor and the reviewer for their insightful comments which led to major improvements in the quality of this manuscript. We thank J. Jemmot for sharing the SAM study (NIH grant 1 R01 HD053270). This work was founded by NIH grants R37 AI 51164, AI 24643, AI113251, ES020337, and AI104459. Portions of this research were conducted on the Cluster at Harvard Medical (NIH grant NCRR 1S10RR028832-01).

References

Belitser, S. V., Martens, E. P., Pestman, W. R., Groenwold, R. H., Boer, A., and Klungel, O. H. (2011). Measuring balance and model selection in propensity score methods. *Pharmacoepidemiology and Drug Safety* **20**, 1115–1129.

Berlin, J. A., Kimmel, S. E., Have, T. R. T., and Sammel, M. D. (1999). An empirical comparison of several clustered data approaches under confounding due to cluster effects in the analysis of complications of coronary angioplasty. *Biometrics* **55**, 470–476.

Beunckens, C., Sotto, C., and Molenberghs, G. (2008). A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. *Computational Statistics & Data Analysis* **52**, 1533–1548.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., and Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology* **163**, 1149–1156.

Brumback, B. A., Dailey, A. B., Brumback, L. C., Livingston, M. D., and He, Z. (2010). Adjusting for confounding by cluster using generalized linear mixed models. *Statistics & Probability Letters* **80**, 1650–1654.

Brumback, B. A. and He, Z. (2011). Adjusting for confounding by neighborhood using complex survey data. *Statistics in Medicine* **30**, 965–972.

Glynn, A. N. and Quinn, K. M. (2010). An introduction to the augmented inverse propensity weighted estimator. *Political Analysis* **18**, 36–56.

Hawkes, M., Sivasivugha, E. S., Ngigi, S. K., Masumbuko, C. K., Brophy, J., and Kibendelwa, Z. T. (2013). HIV and religion in the congo: A mixed-methods study. *Current HIV Research* **11**, 246–253.

Huang, Y. and Leroux, B. (2011). Informative cluster sizes for subcluster-level covariates and weighted generalized estimating equations. *Biometrics* **67**, 843–851.

Hubbard, A. E., Ahern, J., Fleischer, N. L., Van der Laan, M., Lippman, S. A., Jewell, N., et al. (2010). To GEE or not to GEE: Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology* **21**, 467–474.

Hudgens, M. G. and Halloran, M. E. (2012). Toward causal inference with interference. *Journal of the American Statistical Association* **103**, 832–842.

Jemmott III, J. B., Jemmott, L. S., OLeary, A., Ngwane, Z., Icard, L. D., Heeren, G. A., et al. (2014). Cluster-randomized controlled trial of an HIV/sexually transmitted infection risk-reduction intervention for South African men. *American Journal of Public Health* **104**, 467–473.

Kaiser, R., Bunnell, R., Hightower, A., Kim, A. A., Cherutich, P., Mwangi, M., et al. (2011). Factors associated with HIV infection in married or cohabitating couples in Kenya: Results from a nationally representative study. *PLoS ONE* **6**, e17842.

Li, L., Shen, C., Li, X., and Robins, J. M. (2013). On weighting approaches for missing data. *Statistical Methods in Medical Research* **22**, 14–30.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Liu, L. and Hudgens, M. G. (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association* **109**, 288–301.

McDaniel, L. S., Henderson, N. C., and Rathouz, P. J. (2013). Fast pure R implementation of GEE: application of the matrix package. *The R Journal* **5**, 181.

Moore, K. and van der Laan, M. (2009). Increasing power in randomized trials with right censored outcomes through covariate adjustment. *Journal of Biopharmaceutical Statistics* **19**, 1099–1131.

Ogburn, E. L. and VanderWeele, T. J. (2014). Causal diagrams for interference. *Statistical Science* **29**, 559–578.

Paik, M. C. (1997). The generalized estimating equation approach when data are not missing completely at random. *Journal of the American Statistical Association* **92**, 1320–1329.

Pepe, M. S. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics-Simulation and Computation* **23**, 939–951.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, 95–133. New York NY: Springer.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846–866.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* **90**, 106–121.

Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association* **102**, 191–200.

Rotnitzky, A., Robins, J. M., and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of the American Statistical Association* **93**, 1321–1339.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.

SAS (2015). The genmod procedure SAS 12.3. http://support.sas.com/documentation/.

Seaman, S., Pavlou, M., and Copas, A. (2014). Review of methods for handling confounding by cluster and informative cluster size in clustered data. *Statistics in Medicine* **33**, 5371–5387.

Seaman, S. R. and White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research* **22**, 278–295.

Stephens, A. J., Tchetgen Tchetgen, E. J., and DeGruttola, V. D. (2012). Augmented generalized estimating equations for improving efficiency and validity of estimation in cluster randomized trials by leveraging cluster-level and individual-level covariates. *Statistics in Medicine* **31**, 915–930.

Tchetgen Tchetgen, E., Glymour, M., Weuve, J., and Shpitser, I. (2012). Specifying the correlation structure in inverse-probability-weighting estimation for repeated measures. *Epidemiology* **23**, 644–646.

Tchetgen Tchetgen, E. J. and VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research* **21**, 55–75.

Tsiatis, A. A. (2006). Improving efficiency and double robustness with coarsened data. *Semiparametric Theory and Missing Data*, 221–272. New York NY: Springer Science & Business Media.

Tsiatis, A. A., Davidian, M., Zhang, M., and Lu, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. *Statistics in Medicine* **27**, 4658–4677.

Van der Laan, M. J. and Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. New York NY: Springer Science & Business Media.

Vansteelandt, S. (2007). On confounding, prediction and efficiency in the analysis of longitudinal and cross-sectional clustered data. *Scandinavian Journal of Statistics* **34**, 478–498.

Vansteelandt, S., Rotnitzky, A., and Robins, J. (2007). Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika* **94**, 841–860.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121–130.

Zhang, M., Tsiatis, A. A., and Davidian, M. (2008). Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics* **64**, 707–715.

Received July 2015. *Revised February* 2016.
Accepted February 2016.

6.8 Collin et al. 2023 (Int. J. Biostat.) Effect of Interventions against COVID-19

Using a population-based Kalman estimator to model the COVID-19 epidemic in France: estimating associations between disease transmission and non-pharmaceutical interventions
Collin A., Hejblum B., Vignals C., Lehot L., Thiébaut R., Moireau P. and **Prague M.** *International Journal of Biostatistics.* Epub ahead of print - Jan 2023.

This article integrates in the third axis of my research "Outreach to Implementation : Evaluation in Population". It has been written by collaborators from Inria and myself.

I selected this paper as it exemplifies the synergism in research that emerged during the COVID-19 pandemic. The SISTM team and I initially aimed to estimate the impact of NPI on the spread of COVID-19. However, our traditional tools, which required a deterministic temporal link between NPI and virus transmission, proved too limiting. We developed a new methodological tool that employs Population Kalman filters to estimate time-varying transmission rates. This approach allowed us to generate important public health results. The main challenge we faced was in publishing our findings. The fitted model required constant updates to keep pace with evolving data, highlighting a disconnect between the timelines of research, publication and political decision-making.

Annabelle Collin, Boris P. Hejblum, Carole Vignals, Laurent Lehot, Rodolphe Thiébaut, Philippe Moireau and Mélanie Prague*

Using a population-based Kalman estimator to model the COVID-19 epidemic in France: estimating associations between disease transmission and non-pharmaceutical interventions

https://doi.org/10.1515/ijb-2022-0087

Received July 13, 2022; accepted November 8, 2022; published online January 6, 2023

Abstract: In response to the COVID-19 pandemic caused by SARS-CoV-2, governments have adopted a wide range of non-pharmaceutical interventions (NPI). These include stringent measures such as strict lockdowns, closing schools, bars and restaurants, curfews, and barrier gestures such as mask-wearing and social distancing. Deciphering the effectiveness of each NPI is critical to responding to future waves and outbreaks. To this end, we first develop a dynamic model of the French COVID-19 epidemics over a one-year period. We rely on a global extended Susceptible-Infectious-Recovered (SIR) mechanistic model of infection that includes a dynamic transmission rate over time. Multilevel data across French regions are integrated using random effects on the parameters of the mechanistic model, boosting statistical power by multiplying integrated observation series. We estimate the parameters using a new population-based statistical approach based on a Kalman filter, used for the first time in analysing real-world data. We then fit the estimated time-varying transmission rate using a regression model that depends on the NPIs while accounting for vaccination coverage, the occurrence of variants of concern (VoC), and seasonal weather conditions. We show that all NPIs considered have an independent significant association with transmission rates. In addition, we show a strong association between weather conditions that reduces transmission in summer, and we also estimate increased transmissibility of VoC.

Keywords: COVID-19; epidemic modeling; Kalman filters; non-pharmaceutical interventions; population estimation.

1 Introduction

The World Health Organization declared a pandemic COVID-19 on March 11, 2020. This disease is caused by infection with the SARS-CoV-2 virus. By April 30, 2021, more than 150 million cases have been confirmed

***Corresponding author: Mélanie Prague**, ISPED Inserm U1219 Bordeaux Population Health Bureau 23 146 rue Leo Saignat CS 61292 33076 Bordeaux Cedex, France, E-mail: melanie.prague@inria.fr. https://orcid.org/0000-0001-9809-7848

Annabelle Collin, Inria, Inria Bordeaux – Sud-Ouest, Bordeaux INP, IMB UMR 5251, Université Bordeaux, Talence, France, E-mail: annabelle.collin@inria.fr. https://orcid.org/0000-0001-9552-3389

Boris P. Hejblum and Laurent Lehot, Inria, Inria Bordeaux – Sud-Ouest, Talence, Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR 1219, F-33000 Bordeaux, France; and Vaccine Research Institute, F-94000 Créteil, France, E-mail: boris.hejblum@u-bordeaux.fr (B.P. Hejblum), laurent.lehot@u-bordeaux.fr (L. Lehot). https://orcid.org/0000-0003-0646-452X (B.P. Hejblum). https://orcid.org/0000-0001-8562-9550 (L. Lehot)

Carole Vignals and Rodolphe Thiébaut, Inria, Inria Bordeaux – Sud-Ouest, Talence, Univ. Bordeaux, Inserm, Bordeaux Population Health Research Center, SISTM Team, UMR 1219, F-33000 Bordeaux, France; Vaccine Research Institute, F-94000 Créteil, France; and CHU Pellegrin, F-33000 Bordeaux, France, E-mail: carole.vignals@u-bordeaux.fr (C. Vignals), rodolphe.thiebaut@u-bordeaux.fr (R. Thiébaut). https://orcid.org/0000-0002-5235-3962 (R. Thiébaut)

Philippe Moireau, Inria, Inria Saclay-Ile de France, France and LMS, CNRS UMR 7649, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, France, E-mail: philippe.moireau@inria.fr. https://orcid.org/0000-0001-9996-5213

worldwide, including 3.16 million deaths. While the majority of infected cases have a mild form (upper respiratory tract infection symptoms) with no special care needs [1], about 3% of cases, especially the elderly, require hospitalization for treatment, such as oxygen therapy [2–4]. About 17% of these cases are severe forms (severe acute respiratory syndrome) that require treatment in the intensive care units (ICU) and possibly mechanical ventilation [5].

The COVID-19 pandemic has stretched modern healthcare systems around the world to their limits. Because SARS-CoV-2 is an emerging pathogen, the entire human population is vulnerable to infection. Wherever there are outbreaks of SARS-CoV-2, there has been an increase in hospital admissions and especially in the need for intensive care units. Because of an infectious phase that begins before any symptoms are apparent and a significant proportion of a- or pauci-symptomatic infections [6], the spread of SARS-CoV-2 is particularly difficult to control [7]. In response, most governments have taken drastic public health measures, also known as non-pharmaceutical interventions (NPI), to reduce the transmission of SARS-CoV-2 in their populations and thus reduce the pressure on their health systems. In particular, the French government has adopted the concept of a "graduated response" to the pandemic and has deployed an arsenal of different NPIs – some very stringent, others less stringent – in response to the COVID-19 national epidemic situation. Hale et al. [8] have created a stringency index that helps to understand how strong the measures have been over time. However, this indicator does not allow us to distinguish the effectiveness of each NPI, which is critical for future preparedness response plans. Because all NPIs have economic, psychological, and social costs, it is critical to assess their impact on SARS-CoV-2 transmission and on the dynamics of the COVID-19 epidemic.

Many studies relied on mechanistic models of epidemics to either predict progression [9], evaluate vaccine prioritization strategies [10], or retrospectively measure the impact of NPIs. Early in the epidemic, the focus was on the timing of NPI initiation [11, 12] rather than their impact. Disentangling the impact of individual NPIs is a complex problem, in particular because their allocation is not random and depends on the state of the epidemic. Many attempts aggregated data from multiple countries. Some worked with time-series regression based on incidence data [13–16], with semi-mechanistic models and assessed the percentage reduction in the effective reproductive number [17–22] or with advanced machine learning approaches [23]. Instead, we limit ourselves to France, to avoid potential confusion bias in the effect estimation due to differences in behavior and adherence levels across various populations. Most of the work published at the country level focused on a single aggregate NPI such as the Oxford COVID-19 Government Response Tracker [8], a very early epidemic [24, 25], or a very limited number of interventions, see Brauner et al. for a review [26]. Only a few works were interested in multiple waves of epidemics [27–29]. In this study, as in only a few other works as for example Ge et al. [30], we aim to consider a rather long observation period (the first 12 months of the pandemic).

In France, although some studies have quantified the effect of several NPIs [3, 31], the authors base their results on the earliest phase of the epidemic and examine only a limited number of NPIs. Our work focuses instead on the impact of individual NPIs on the transmission rate. This is a better and more valid indicator than direct epidemic curves or the reproductive number, because it is independent of the proportion of the population infected. Finally, previous work does not take into account the prevalence of vaccination, the spread of new variants, and the importance of weather in estimating the impact of NPIs. Quantifying the associations between NPIs and transmission rate leads to a challenging estimation problem, including concerns about the practical identifiability of each effect.

In this work, we propose a two-step approach. First, we estimate the transmission rate of SARS-CoV-2 and its variations in the 12 non-island French regions over a period of more than one year – from March 2, 2020 to March 28, 2021. Second, we use linear regression to estimate the associations between different NPIs and transmission rates. We account for seasonal weather conditions during the pandemic, as well as the occurrence of non-historical variants of concern (VoC) and an increasing proportion of vaccinated individuals. This methodology is very innovative in itself, and has not been applied to other datasets before. It involves the application of a new population-based Kalman filter estimate (proposed by the authors in Collin et al. [32]), which extends classical Kalman filters to the parameters estimation across a whole population

of subjects. It is to note that association of NPIs with epidemic dynamics has never been formally studied using French data. Finally, it is also noteworthy that placing the estimation of epidemic dynamics into a population approach is singular. More precisely regarding the inference procedure, the first step is to estimate transmission rates in the 12 mainland French regions. This is a major challenge because the available data are very sparse and noisy, and the parametric form of transmission rates is also unknown. By assimilating data across multiple geographic regions and coupling public data with a dynamic mechanistic model, smooth transmission rates can be estimated using a Kalman filter approach [33, 34] – as already used in epidemiology for COVID-19 spread [35–37] or for other epidemics with regional variability [38]. More specifically, we develop a sophisticated method to address this difficult problem based on two important methodological innovations: (1) in the model with the introduction of time-varying dynamics for the transmission rate, including a Wiener process that accounts for modeling errors, and (2) in the way the population is integrated, as we use a new method from the Kalman filter that is compatible with population approaches. This method, in which the log-likelihood function is – estimated by, for example, the unscented Kalman filter [39] – elegantly couples data across multiple geographic regions, is presented in Collin et al. [32]. These two innovations are coupled in a strategy that allows estimation of smooth transmission rates without imposing prior-knowledge on their shape or evolution, and relate them to key NPIs, seasonal weather conditions, vaccination coverage, as well as new VoC circulation.

Section Materials and methods presents the supporting data and the underlying mechanistic model for the epidemic, before introducing our two-step strategy to estimate first transmission rates and their associations with NPIs using a multivariate regression approach. Section Results highlights the different effects estimated for the NPIs implemented in France during the first year of the pandemic. Finally the strengths and limitations of our approach are acknowledged in Section Discussion.

2 Materials and methods

Open data on the French COVID-19 epidemic, including data on hospitalizations, NPI implementations, VoC prevalence, and the vaccination program are presented below. We then introduce our dynamic modeling for the COVID epidemic, relying on an extended SIR type model. Finally, we describe our strategy for estimating transmission rates using a population-based Kalman filter approach from Collin et al. [32] to study the associations between transmission rates and NPIs, seasonal weather conditions, and VoC.

2.1 Available data

2.1.1 Hospitalization data: Hospitalization data come from the SI-VIC database (Système d'Information pour le suivi des VICtimes), a government system established in 2016 to identify and track victims in exceptional circumstances (e.g., terrorist attacks). Since beginning of March 2020, the SI-VIC database of the French public health national agency (Santé Publique France) provides the daily number of hospitalized COVID-19 patients, at different geographical levels. In this work, we focus on the 12 non-insular regions of mainland France.

Each entry in the SI-VIC database indicates a patient hospitalized in connection with COVID-19. To qualify, at least one of two criteria must be met by this patient: (i) a biologically confirmed diagnosis of COVID-19 (e.g., RT-PCR positive test result) or (ii) a chest examination CT scan suggestive of COVID-19. In this analysis, we rely on the daily incident number of hospitalizations (denoted $Y^{H_{\text{in}}}$) and the total prevalence of persons hospitalized daily (denoted Y^{H}, corresponding to the number of occupied hospital beds). We consider a period of 391 days (from March 2, 2020 to March 28, 2021). To compare the magnitude of the epidemic in each region, we standardized the data using the population size in each region to present hospitalizations per 100,000 inhabitants. The two data series are displayed in Appendix A, Figure 8.

2.1.2 Non-pharmaceutical interventions (NPIs): The timing and modalities of the various NPIs implemented in France during the epidemic are taken from the French government's website summarizing the measures [40]. In France, public health interventions were multifaceted. In our analysis, we considered the following summary NPI that took place during the first year of the epidemic in France: (i) first lockdown (with two phases of easing/reopening, as described below), (ii) second lockdown (with one phase of easing/reopening, as described below), (iii) curfew from 8 PM, (iv) curfew from 6 PM, (v) closure of schools, (vi) closure of bars and restaurants, (vii) barrier gestures (including all mandatory hygiene protocols: physical distancing, hand

washing, part-time distancing, remote work and wearing masks in public spaces). Of note, NPIs such as travel bans, enhanced testing, contact-trace isolation were ignored to ensure identifiability because they were either of difficult-to-quantify magnitude, or enforced in complete overlap with other NPIs. In addition, partial interventions at the sub-regional level were not taken into account. This resulted in a fairly similar profile of interventions across regions, as most of them were applied simultaneously in the 12 regions of interest. Figure 1 provides an overview of all 10 NPIs considered over time in the *Île-de-France* region. The timing of the introduction of the other NPIs is quite similar in the other regions, differing only by a few days, (1) due to the school vacations (one week postponement) and (2) due to the earlier introduction of the 8 pm curfew in the areas most affected by COVID-19, see Figure 9 in Appendix A for more details.

The first and second lockdowns were differentiated because they had distinct modalities resulting in different behaviors and thus potentially impacting transmission differently. For example, during the first lockdown (from March 17, 2020, to May 11, 2020), the entire population was required to work from home - the only exceptions were critical workers from the medical sector, food industry or security, while personal outings could not exceed 1 h within a 1 km radius from home. During the second lockdown (from October 29, 2020, to December 15, 2020), on-site work was allowed if working from home was not possible, and outings were limited to 3 h within a 20 km radius. In addition, the end of the initial lockdown was gradually divided by the government into three official phases (Phase 1: May 11, 2020 to June 2, 2020, Phase 2: June 2, 2020 to June 22, 2020, and Phase 3: after June 22, 2020) with many evolving measures such as authorized distance of travel, reopening of cultural sites (i.e., museums), reopening of non-essential stores, etc. A government campaign to raise awareness of barrier gestures began at the end of the first lockdown. Masks and hand washing were mandatory in many places such as public transportation, schools, and businesses. We thus presume that the *barrier gestures* NPI began on May 11, 2020, assuming that most mandatory measures with potentially high impact have been implemented by that time. Regarding the second lockdown, non-essential stores reopened 2 weeks before the end of the lockdown before Christmas. To account for this, we divide the second lockdown into two phases: a full lockdown until November 28, 2020, and a reduced lockdown thereafter.

School closures were documented by the official holiday schedule, which could vary from region to region. In France, schools were also closed during the first lockdown. In addition, from May 11, 2020 (the end of the first lockdown) to July 4, 2020 (the end of the school year), schools gradually reopened and student enrollment slowly increased back to normal. Estimated student

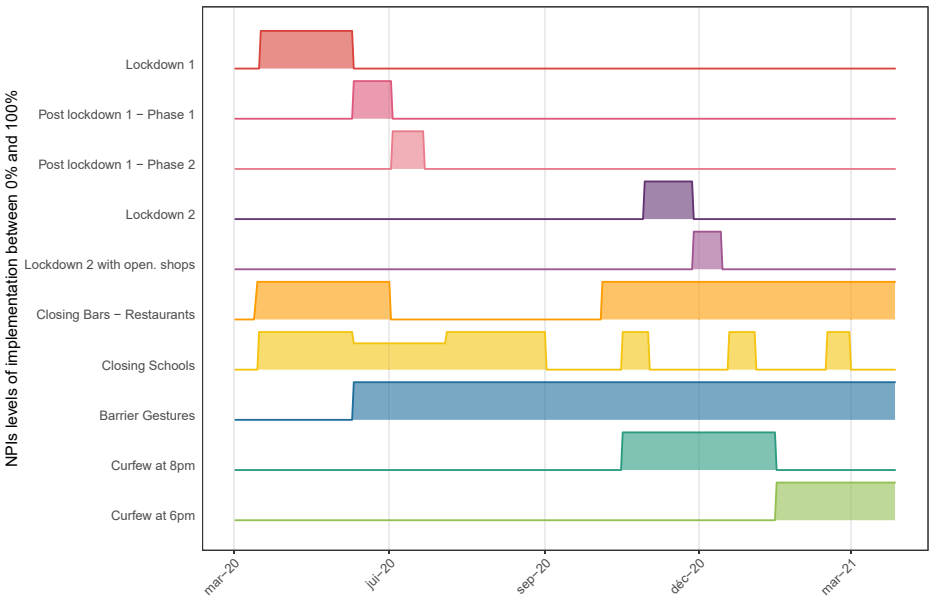


Figure 1: Implementation of major NPIs in region "Ile de France". The timing of NPIs implementation is fairly similar across all regions and only differs for a few days in the implementation of school closing because of holidays and curfews 8 pm (which was initiated few days earlier in regions with highest COVID-19 burden). All NPI are on/off except school closure right after the first lockdown due to the implemented lock down lifting process as described in the text.

enrollment averaged 30% of school capacity during this transition period according to general media coverage. Full closures of all bars and restaurants occurred twice during the study period: first a few days before the implementation of the first lockdown in March 2020 (as with school closures) and at the beginning of the second lockdown in October 2020. After the end of the first lockdown in June 2020, this measure was gradually lifted in all regions. At the end of the second lockdown, the measure was not lifted and was still in effect at the end of the study period.

Curfew was first enforced on October 17, 2020, in the 12 regions of interest in several large cities and in *Île-de-France* from 9 PM to 6 AM. It was extended to 54 departments (sub-regional administrative units) on October 22. During the second lockdown from Oct. 30 to Dec. 15, 2020, a national curfew from 8 PM to 6 AM was en forced. From January 2 to 12, 2021, curfew start at the department level was gradually advanced to 6 PM until January 16, when it became 6 PM for all the 12 regions of interest. Finally, on March 20, 2021, curfew starting time was changed back to 7 PM nationally. For the sake of simplicity, we considered only two different NPIs, grouping together curfews starting at 9 PM or 8 PM on the one hand, and those starting at 6 PM or 7 PM on the other hand, and we considered a region under curfew only if it was enforced in the entire region. Assuming that curfews and lockdowns induce different behaviors, curfews are not considered to be included in lockdowns, even though it is forbidden to go out in the evening during lockdowns. Similarly, 6 PM curfew and 8 PM curfew were considered distinct interventions rather than nested interventions because they are likely to produce different behaviors, and thus have different impact on transmission.

2.2 Other exogenous variables: weather, Variants of Concern (VoC), vaccination coverage

2.2.1 Weather conditions: The role of weather conditions in SARS-CoV-2 transmission remains controversial, and early publications have been criticized for their inconsistent results [41]. Nonetheless, based on comparisons with other respiratory infections, the potential effects of temperature and humidity on aerosolized and fomite transmission pathways are based on sound mechanistic arguments [42]. As the Northern Hemisphere underwent a second winter season during the pandemic, evidence linking transmission to seasonal trends in temperature and humidity (also affecting human behavior and indoor/outdoor gathering) appears more robust [42–45]. Daily weather data – namely temperature in Celsius (T), relative humidity in percent (RH), and absolute humidity in g m^{-3} (AH) – measured from meteorological stations were extracted from the National Oceanic and Atmospheric Administration database using the R package *worldmet*. All stations located in the region or within 10 km of the region boundary were used to compute daily regional weighted averages (to account for differences in population density within a region, we applied a weighting based on the population within a 10-km radius around each station giving more weight to weather conditions in densely populated areas). We use the Index PREDICT of Transmissibility of COVID-19 (IPTCC), as defined by Roumagnac et al. [46]:

$$\mathrm{IPTCC} = 100\, \mathrm{e}^{-\frac{1}{2}\left[\frac{(T-7.5)^2}{196}+\frac{(RH-75)^2}{625}+\frac{(AH-6)^2}{2.89}\right]}.$$

This IPTCC index ranges from 0% to 100%, the smaller the less favorable the conditions for COVID-19 transmission. In France, we observe a seasonal variation of IPTCC low in summer and higher otherwise, with a northeast/southwest gradient. For easier interpretation, we normalized this IPTCC (forcing its range to one), subtracting the national average, and inverting it (see grey curves in Figure 2). Thus, the lower the value, the closer the temperature and humidity conditions were to the optimal transmission conditions defined by Roumagnac et al. [46]. Finally, to focus on seasonal variations, a *loess* smoothing with a span of 0.2 was applied, resulting in a smooth weather variable denoted W below displayed in Figure 2 for each region. Summer (with higher values of this weather variable) is clearly standing out from winter (with lower values). Taking $W = 0$, we consider the global average value over all French regions during the study period, and $W_i(t)$ denotes the weather in the ith region at date t.

2.2.2 SARS-CoV-2 variants of concern (VoC): Certain variants of the SARS-CoV-2 virus have been classified as VoC by national and international health authorities because they affect transmissibility or virulence or reduce the effectiveness of interventions [47]. Within our study period, French health authorities have conducted surveys to estimate the prevalence of three VoCs: 20I/501Y.V1 (alpha), 20H/501Y.V2 (beta), and 20J/501Y.V3 (gamma). The delta and omicron lineage VoC surveys took place after this study. We therefore included the cumulative proportion of cases infected by any of these three first VoC as a possible covariate associated with transmissiona and we denote this variable as $\mathrm{VoC}_i(t)$ in the following.

We used data from two cross-sectional so-called "flash" surveys conducted on January 7 & 8, 2021, and on January 27, 2021 [48, 49], as well as the weekly estimate of VoC spread provided by the SI-DEP database at the regional level from February 12, 2021, to March 28, 2021 [50]. Between January 8, 2021, and February 12, 2021, the estimated proportion of the sum of the three VoC increased from a national average of 3.3% to a national average of 46%. To fill in the missing data, we assume that the proportion is equal 0% before January 8, 2021, and that the trend is linear between January 8 and January 27, 2021, and linear between January 27 and February 12, 2021. It should be noted that a logistic and an exponential growth were also investigated, without resulting in any significant change in our conclusions (results not shown). Since no data were reported for the *Bourgogne-Franche-Comté* region on January 27, 2021, only one slope was estimated from January 8 to February 12. VoC fraction in each region is displayed over time in Appendix A.

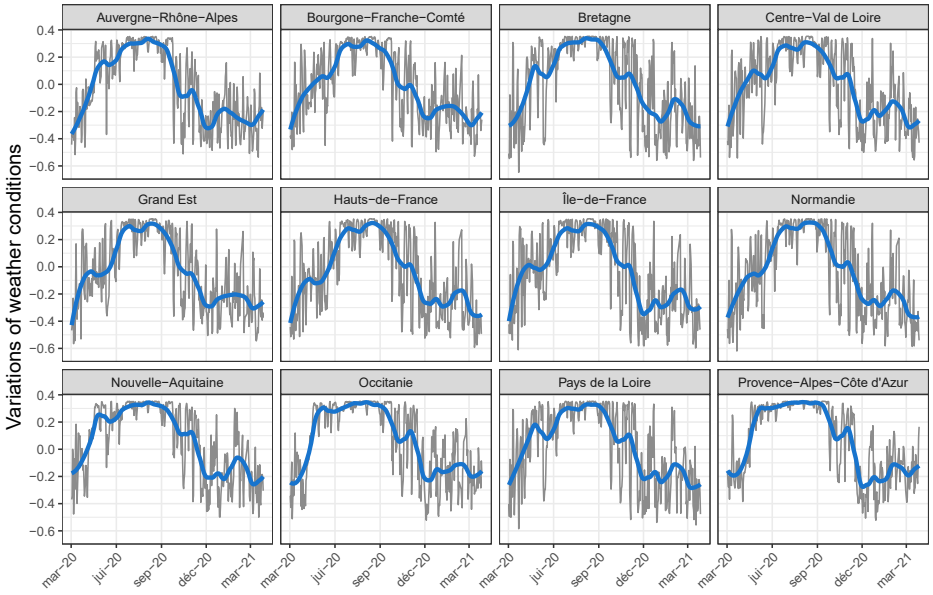


Figure 2: Weather variable modeling the seasonal weather conditions of the 12 regions of interest in grey and after smoothing in blue (denoted W). The higher the value, the lower the favorable conditions (temperature and humidity) for transmission of COVID-19.

2.2.3 Early vaccination: Vaccination began in France on December 27, 2020, at which time three COVID vaccines were licensed and available: BNT162b2 mRNA (Pfizer), ChAdOx1 nCoV-19 (AztraZeneca), and mRNA-1273 (Moderna). To account for the starting vaccination campaign, we used the database VAC-SI [51], which records the cumulative percentage of the population vaccinated with at least one dose of vaccine over time. Vaccination was initially prioritized for the elderly aged 75 years and older. The proportion of the total population vaccinated increased to approximately 12% by the end of the study period, and regional population vaccination coverage rates over time are shown in Appendix A.

2.3 Modelling the epidemic

2.3.1 The mechanistic model: We model the evolution of the COVID-19 epidemic using an extended SEIR model [52, 53], called the SEIRAH model, in which the population of size N is divided into 5 compartments: susceptible S, latently exposed E, symptomatically infectious I, asymptomatic/pauci-symptomatically infectious A, hospitalized H, removed R (i.e., both recovered and deceased), see Figure 3. The number of vaccinated individuals denoted by V, is assumed to be known, see Section 2.2.3. The dynamics of such a model is given by

$$\begin{cases} \dot{S} = -\underline{b}\left(1-\frac{V}{N}\right)\frac{S(I+\alpha A)}{N} \\ \dot{E} = \underline{b}\left(1-\frac{V}{N}\right)\frac{S(I+\alpha A)}{N} - \frac{E}{D_E} \\ \dot{I} = \frac{r_E}{D_E}E - \frac{1-r_I}{D_Q}I - \frac{r_I}{D_I}I, \\ \dot{R} = \frac{r_I I + A}{D_I} + \frac{H}{D_H} \\ \dot{A} = \frac{1-r_E}{D_E}E - \frac{A}{D_I} \\ \dot{H} = \frac{1-r_I}{D_Q}I - \frac{H}{D_H} \end{cases} \tag{1}$$

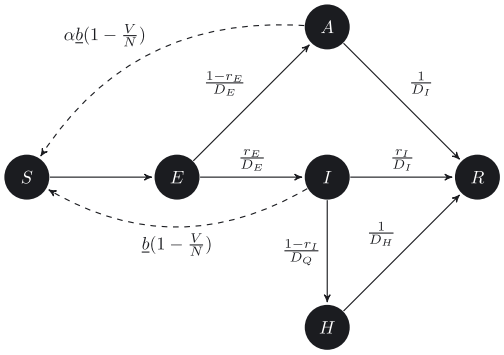


Figure 3: SEIRAH model representation.

where $\alpha, r_E, D_E, r_I, D_I, D_Q, D_H$ are time-independent parameters described in Table 1 while $\underline{b}$ is a function of time modeling the disease transmission rate.

2.3.2 The observation model: The two quantities Y^H and $Y^{H_{\text{in}}}$ relate to the solutions of System (1) respectively as for all regions $i = 1, \ldots, 12$, for all observation time in days $j = 1, \ldots, 391$: $Y^H_{ij} = H_i(j) + \epsilon^H_{ij}$ and $Y^{H_{\text{in}}}_{ij} = \frac{(1-r_I)}{D_Q} I_i(j) + \epsilon^{H_{\text{in}}}_{ij}$, in which ϵ^H_{ij} and $\epsilon^{H_{\text{in}}}_{ij}$ represent normally distributed constant measurement errors.

2.3.3 Effective reproductive number and attack rates: When individuals are homogeneous and mix uniformly, the effective reproductive ratio $R_{\text{eff}}(t)$ is defined as the mean number of infections generated during the infection period of a single infectious case at time t. In this model, the effective reproductive ratio can be written as a function of model parameters (see Appendix C for details):

$$R_{\text{eff}}(t) = \underline{b}(t)\Big(1 - \frac{V}{N}\Big)\frac{S(t)}{N}\left(D_I\alpha(1-r_E) + \frac{D_I D_Q r_E}{(1-r_I)D_I + r_I D_Q}\right). \tag{2}$$

When neglecting the deaths, the proportion of infected individuals assuming no waning immunity – also called attack rates – among the population in each region at a given date is given by:

$$\frac{E+I+R+A+H}{N} = 1 - \frac{S}{N}.$$

Table 1: Model parameters for the SEIRAH model and associated values.

Parameter	Interpretation	Value
$\underline{b}$	Transmission rate of infectious cases	Region specific – estimated
r_E	Ascertainment rate	0.844 [54]
r_I	Non hospitalized rate	0.966 [4]
α	Ratio of transmission between A and I	0.55 [55]
D_E	Latent (incubation) period (days)	5.1 [56]
D_I	Infectious period (days)	5 [57]
D_Q	Duration from I onset to H (days)	$11 - D_E = 5.9$ [58]
D_H	Hospitalization period (days)	18.3*
N	Population size	Region specific

*Computed using the correlation between the data $Y^{H_{\text{in}}}$ and Y^H when considering region data, see Appendix B.

2.4 A population-based Kalman filter to estimate the transmission rate

2.4.1 Framing the problem: To support the modeling decisions underlying the disease transmission rate $\underline{b}$, and more generally the inference procedure chosen, we applied Kalman-based estimation strategies [33, 34] to the epidemic model (1). We define the global transmission rate $t \mapsto b(t)$, which accounts for the proportion of susceptibles removed from the system by vaccination:

$$b(t) \overset{\text{def}}{=} \underline{b}(t)\left(1 - \frac{V(t)}{N}\right).$$

We propose to estimate b and finally $\underline{b}$ using an estimate of vaccinated individuals V (see Subsection 2.2.3). In each region i, we then introduce a dynamic equation for b of the form

$$\mathrm{d}b^i(t) = g^i(t)\mathrm{d}t + \mathrm{d}v^i(t), \tag{3}$$

where v^i consists of a Wiener process such that for all $t, s \geq 0$, $v^i(t) - v^i(s) \sim \mathcal{N}(0, (t-s)\sigma_v)$ with σ_v^2 known and constant across all regions, and g^i a function describing the evolution of global transmission rates b^i in each region.

After discretization using Euler's time scheme with a sufficiently small time step δt, we obtain a discrete-time dynamical system applied to the variable $x = (E, I, R, A, H)^\intercal \in \mathbb{R}^5$, for each region $i = 1, \ldots, 12$:

$$z^i_{n+1} \overset{\text{def}}{=} \begin{pmatrix} x^i_{n+1} \\ b^i_{n+1} \\ \theta^i_{n+1} \end{pmatrix} = \begin{pmatrix} x^i_n + \delta t\, f(x^i_n, b^i_n, \theta^i_n) \\ b^i_n + \delta t\, g^i(t_n, \theta^i_n) \\ \theta^i_n \end{pmatrix} + \begin{pmatrix} 0_5 \\ 1 \\ 0_{N_p} \end{pmatrix} v^i_n, \tag{4}$$

where f accounts for the dynamics of (E, I, R, A, H) in (1), while S is subsequently reconstructed as $S = N - (E + I + R + A + H)$ in each region. In the discrete-time system, $\left(v^i_n\right)_{n\geq 0}$ now represent independent random variables, normally distributed with 0 mean and variance equal to $\sigma_v^2 \delta t$. Moreover, if constant parameters need to be estimated, the vector $\theta \in \mathbb{R}^{N_p}$ summarises all these parameters. For the estimation below, we transform the variable z to account for biological constraints. Since all state variables are positive and bounded by N, the total population size of the region, we transform the state variable with $x \mapsto \mathrm{logit}(x/N)$. We also apply a similar transformation to $b \mapsto \mathrm{logit}(b/\max_b)$. Note that the state and transmission rate variables in the transform space tend to have Gaussian distributions. After calibration, we set $\max_b = 1.5$ and verified that other values did not significantly change the results (result not shown).

2.4.2 Population approach: To perform the estimation, we rely on an extension of the classical Unscented Kalman Filter (UKF) [33, 39, 59]. The peculiarity of this application is that multiple data series in multiple regions are observed together, since we observe multiple realizations of the same epidemic (each region being a different realization). To account for parameter correlation across the different regions in our Kalman estimation, we follow a recently proposed population-based Kalman formulation [32]. As in mixed effects models [60], each initial uncertainty variable z^i_0 is assumed to be randomly distributed around a common population intercept z^{pop}_0 with a Gaussian distribution of unknown covariance $\boldsymbol{Q}_0$, namely:

$$z^i_0 \sim_{\text{i.i.d.}} \mathcal{N}\left(z^{\text{pop}}_0, \boldsymbol{Q}_0\right).$$

By treating the population intercept as an empirical mean over the population members, we obtain a classical filtering problem [33] on the aggregate variable $\mathbf{z} = \left(z^1, \ldots, z^{N_r}\right)^\intercal$ when constructing our objective function. The only difference is the formulation of the initial covariance prior $\hat{\boldsymbol{P}}_0$, which couples observations across regions and can be written as follows:

$$\hat{\boldsymbol{P}}_0^{-1} = \frac{1}{N_r^2}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\begin{pmatrix} 1 & \ldots & 1 \end{pmatrix} \otimes \boldsymbol{M} + \left[\mathbb{1}_{N_r} - \frac{1}{N_r}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\begin{pmatrix} 1 & \ldots & 1 \end{pmatrix}\right] \otimes \hat{\boldsymbol{Q}}_0^{-1},$$

where $\otimes$ denotes a Kronecker product, $\hat{\boldsymbol{Q}}_0$ is a prior of $\boldsymbol{Q}_0$, and $\boldsymbol{M}$ is a small penalization matrix guaranteeing the overall invertibility of $\hat{\boldsymbol{P}}_0$. As a result, the matrix $\hat{\boldsymbol{P}}_0$ is non-block diagonal with respect to the region i and thus all the dynamics of the regions are coupled. The resulting discrete-time Kalman estimator ties all regions together to obtain a population-based estimate. Note that in such a strategy, it is possible to force a variable in the population to be constant by simply choosing $\hat{\boldsymbol{Q}}_0$ such that $\mathrm{Tr}\left(\hat{\boldsymbol{Q}}_0^{-1}\right)$ is very small with respect to $\mathrm{Tr}(\boldsymbol{M})$. Conversely, a small $\mathrm{Tr}(\boldsymbol{M})$ with respect to $\mathrm{Tr}\left(\hat{\boldsymbol{Q}}_0^{-1}\right)$ in a large population $N_r \gg 1$ will encourage regions to remain independent from each other. Given prior knowledge, our Kalman implementation uses the available measurements $\left(Y^H_{ij}, Y^{H_{\text{in}}}_{ij}\right)_{1\leq i\leq 12, 0\leq j\leq 391}$ to recursively compute the following estimates:

$$\hat{\boldsymbol{z}}_n \simeq \mathbb{E}(\boldsymbol{z}_n | \left(Y_{ij}^{H}, Y_{ij}^{H_{\text{in}}} \right), 1 \leq i \leq 12, 1 \leq j \leq n), \quad 0 \leq n \leq 391,$$

and

$$\hat{\boldsymbol{P}}_n \simeq \mathbb{C}\text{ov}(\boldsymbol{z}_n - \hat{\boldsymbol{z}}_n | \left(Y_{ij}^{H}, Y_{ij}^{H_{\text{in}}} \right), 1 \leq i \leq 12, 1 \leq j \leq n), \quad 0 \leq n \leq 391.$$

Taking advantage of $\hat{\boldsymbol{z}}_n$ gathering the augmented state $\left(x_{n+1}^i, b_{n+1}^i, \theta_{n+1}^i\right), 1 \leq i \leq 12$ of the 12 regions, a simple post-processing over the regions yields estimates of b and the state variables. The fact that the state variables S, E, I, R, A, and H are included in the augmented state means that the errors in the initial conditions are corrected over time, allowing for different prevalences in different regions. We refer the reader to Collin et al. [32] for more details on this Kalman-based population approach.

2.4.3 Estimation strategy: Since we want to inject as little information as possible about the shape of the transmission rate b, we will first assume that the Wiener process ν defined in Eq. (3) is a time-dependent function, thus encompassing the entire dynamics of b. This is consistent with setting $g \equiv 0$ (indicating no prior knowledge of the relationships between the evolution of the transmission rates and the NPIs). However, to avoid overfitting, our goal is to distinguish the latent trajectory of b from noise in the observations. We use a 3-step approach for smoothing the trajectories of b, described below:

1. Estimate an appropriate prior for the initial transmission rates b before the start of any NPI. We use data before the first lockdown (10 days available) and assume that the transmission rate $b^i(t) = b^i_{\text{init}}$ is a constant for $t = 1, \dots, 10$ days. In other words: We apply the population Kalman filter estimate described above, with θ from Eq. (4) reduced to b_{init}.
2. Estimate the shape of $b(t)$ with a prior for the initial value, but without a prior for the dynamics. We set the initial value of b^i for times $t = 1, \dots, 10$ days to b^i_{init} and take $g(t) \equiv 0$ such that Eq. (3) rewrites $\mathrm{d}b(t) = \mathrm{d}\nu(t)$. We apply the population Kalman filter estimation described above. Note that the parameter vector θ from Eq. (4) is now empty, and since there is no information about b, the model error is very large and could lead to overfitting. We then create a prior for the dynamics of b by fitting a parametric form based on the sum of logistic functions (well suited to modeling observed variations such as stiff lockdowns or smooth unlocks) to the weighted average trajectories of b over all regions using the least squares method. The number of logistic functions summed is based on the observed number of principal changes in the variations of the weighted average trajectories of b. This now gives us a prior g for the dynamics.
3. Estimate the shape of b with a prior for the initial value and an informative prior for the dynamics. Finally we set the initial value of b^i for times $t = 1 \dots 10$ days to b^i_{init} and take $\mathrm{d}b(t) = g(t)\mathrm{d}t + \mathrm{d}\nu(t)$ as in Eq. (3). We apply the population Kalman filter estimation described above. Since the dynamics of b still contains modeling noise, the shape of b differs from the prior transmission rate defined in Step 2. This final estimate will be used to further describe the relationships between transmission rates and NPIs.

2.5 Explanatory model for the transmission rate

2.5.1 Mixed effects model: Using the b estimate obtained from the population-based Kalman filter, in the third step described above, we can determine the association between NPIs, seasonal weather conditions, VoC proportion and the transmission rate $\underline{b} = \frac{b}{(1-V/N)}$. We followed Flaxman et al. [19] and considered interventions to have multiplicative effects. Therefore, we applied a linear mixed effects model to the log transformation of transmission. The model equations are given in Appendix F. It consists of an intercept (representing the average transmission of COVID-19 over all regions without NPIs, without VoC circulation, and for average French weather conditions ($W = 0$)), the effects of the 10 NPI described in Section 2.1.2, the effect of weather, and the effect of VoC fraction. Because the transmission of SARS-CoV-2 is different indoors and outdoors [61], we also added an interaction effect between bar and restaurant closure and weather that accounts for the opening of outdoor seating areas. This is the only interaction included (and studied) to avoid overfitting. Finally, we added random effects to account for heterogeneity between regions. We added a random intercept and random slopes for the effect of the first and second lockdowns as well as the 6 PM curfew because their effects can vary greatly between regions. We assume a full covariance matrix for the random effects, so that the associations in each region may be correlated – in particular, they could be influenced by several factors not accounted for in the epidemic model, such as population density, age distribution, or urbanization.

Note that we have delayed the lockdowns by 7 days. This can be interpreted as a necessary period of time to allow people to organize and adapt (introduction of home-based work, child care, etc.). This decision was motivated by the observed 7-day delay in immediate transmission rates decline obtained with Kalman filters, see 4. To facilitate interpretation of the estimated associations, the parameters were transformed and expressed as percent decrease or increase in transmission by applying the function $x \mapsto 100(\mathrm{e}^x - 1)$. Classical 95% confidence intervals were obtained as $100(\mathrm{e}^{x \pm 1.96\ \mathrm{SE}(x)} - 1)$ assuming normality, where SE is the standard error obtained from the regression.

2.5.2 Interpretation of the association between seasonal weather conditions and transmission rates: The variable for seasonal weather conditions being unitless, so interpretation of its estimated effect is done in comparison to the average weather

conditions in France. To further facilitate understanding, we computed the average relationships between NPIs and transmission rates during both summer and winter periods respectively, using the summer (or winter) average of $W(t)$ from June 21st, 2020 to October 21st, 2020 (or before June 21st, 2020 and after October 21st, 2020) across all regions.

2.5.3 Basic reproductive number: The intercept of the above regression represents the mean transmission rate over all regions when there is no NPI and no VoC and the weather conditions are assumed to be the average weather conditions over a year in France. Thus, substituting it into Eq. (2), directly provides the basic reproductive number and 95% confidence intervals can be calculated using the standard error of these parameters.

3 Results

3.1 Estimation of the transmission rate using a population-based Kalman filter

Step 1 of the estimation provided the initial values for the transmission rate with fairly similar values across regions (average 0.78 sd 0.012, see details in Appendix D). Although higher values were found in regions where the first wave was stronger, the low variability between regions suggests that the magnitude of the first wave is not solely due to a higher transmission rate, but also to differences in epidemic baseline conditions (i.e., the number of exposed and infectious cases, linked to differences in the timing of virus introduction or the occurrence of super-spreading events [62]) and weather conditions. Transmission rates estimated in Step 2 without knowledge of their shape are shown in Figure 4 (top, left) for all regions. Since there is no prior information about b inputted, the model error is very important and leads to over-fitting of the data. In particular, many regions exhibits weekly oscillations related to the under-reporting during weekends. We approximated the weighted average trajectory of b by a sum of 7 logistic functions is, as shown in Figure 4 (bottom, left) which becomes our prior in Step 3. Finally, in Step 3, smooth regional transmission rates b_i are estimated, see Figure 4 (top, right). Since the dynamics still contain modeling noise, the shapes of regional b_i are naturally different from the prior defined in Step 2 (while still smooth).

3.2 Basic and Effective reproductive number

Figure 4 (bottom, right) presents the estimated effective reproductive ratio R_{eff}, with starting values ranging between 3.5 and 4 in all regions. This variability could be due in part to winter weather conditions in early March 2020, and after adjusting for weather condition we estimate the national average basic reproductive number at 3.10 [2.95; 3.26]. Variations of R_{eff} over time shows that it quickly falls below the critical value of 1 after initiation of the first and second lockdowns.

3.3 Attack rates

The attack rate (defined in Section 2.3.3) provides additional knowledge about the number of possible hidden/unmeasured cases. Figure 5 displays the attack rate at several key points in time: at the end of the first lockdown (May 11, 2020), on October 5, 2020, and at the end of our study period (March 28, 2021), while we estimate national attack rates at these time points to be 5.7%, 8.8%, and 25.3%, respectively.

3.4 Associations between NPIs and transmission rates

Table 2 summarizes estimated associations between the NPIs and the transmission rate (from the model introduced in Section 2.5 – residuals and detailed model coefficients can be found in Appendix F). Figure 6 shows the corresponding regional fits. Of note, the values for both curfews are very close (near 30%), and not statistically different (estimated difference is 3% [−2%; 8%]). We show that all NPIs in this analysis reduce transmission and have an effect that is significantly different from zero.

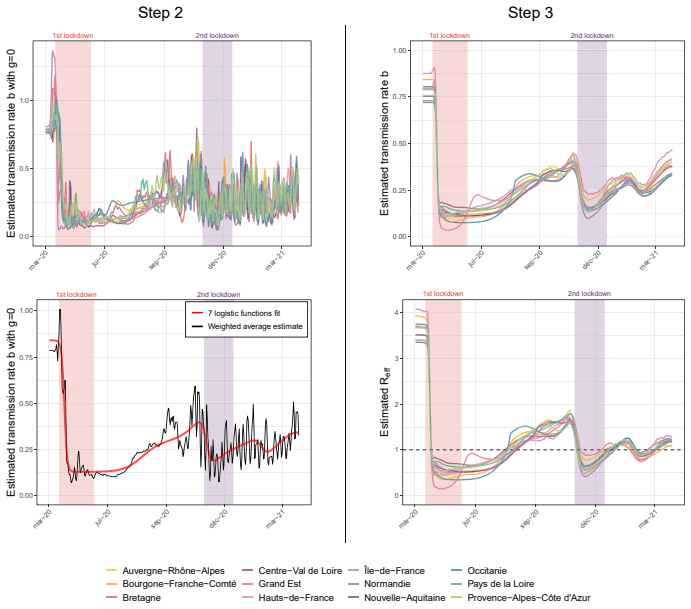


Figure 4: First column (step 2) Top: regional dynamics of b with $g = 0$ (without *a priori*). Bottom: mean value of b over time (black line) obtained in step 2, fitted with 7 logistic functions (dashed red line). **Second column (step 3)** Top: estimated regional dynamics evolution of b with g the 7 logistic functions obtained in step 2 (with $v \neq 0$). Bottom: estimated regional dynamics of R_{eff} with g the 7 logistic functions.

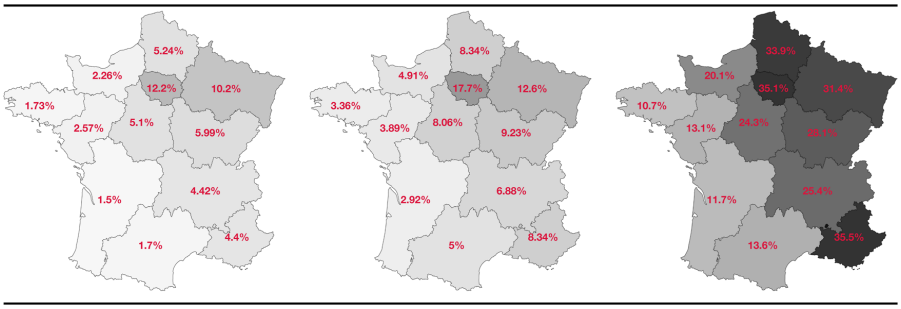


Figure 5: Model estimation for the proportion of naturally immunized individuals in the population (deaths and vaccinated people not taken into account) on May 11th, 2020 (left), on Oct. 5th, 2020 (middle) and on March 28th, 2021 (right).

3.5 Effect of weather

We estimate that, on average, transmission is significantly increased by 10% during the winter period and significantly decreased by 22% during the summer period (compared to the transmission under average weather conditions). Figure 7 (top) shows the estimated impact of seasonal weather conditions on transmission during the study period in each region. The estimated interaction between bar and restaurant closures and

Table 2: Estimation and 95% confidence intervals of the associations between the transmission rate and seasonal weather conditions, VoC proportion, and NPIs. Model AIC = −1388.

Covariate	Modification of the transmission rate	
	Scale	Increase/decrease
NPIs		
Lockdown 1 – delay of 7 days	−78% [−82%; −74%]	↘
Post lockdown 1 – phase 1	−54% [−56%; −52%]	↘
Post lockdown 1 – phase 2	−48% [−50%; −47%]	↘
Lockdown 2 – delay of 7 days	−54% [−57%; −49%]	↘
Lockdown 2 with opened shops	−51% [−53%; −49%]	↘
Closing schools	−7% [−8%; −5%]	↘
Barrier gestures	−46% [−48%; −44%]	↘
Curfew at 6 PM	−30% [−33%; −26%]	↘
Curfew at 8 PM	−28% [−31%; −25%]	↘
Bar & restaurant closure (ref. $W = 0$)	−10% [−13%; −8%]	↘
Bar & restaurant closure, summer (vs. ref.)	−8% [−11%; −4%]	↘
Winter (vs. ref.)	−11% [−14%; −8%]	↘
Other factors		
100% of VoC circulating	22% [15%; 28%]	↗
Weather effect during summer (ref. $W = 0$)	−22% [−24%; −21%]	↘
During winter (ref. $W = 0$)	10% [9%; 11%]	↗

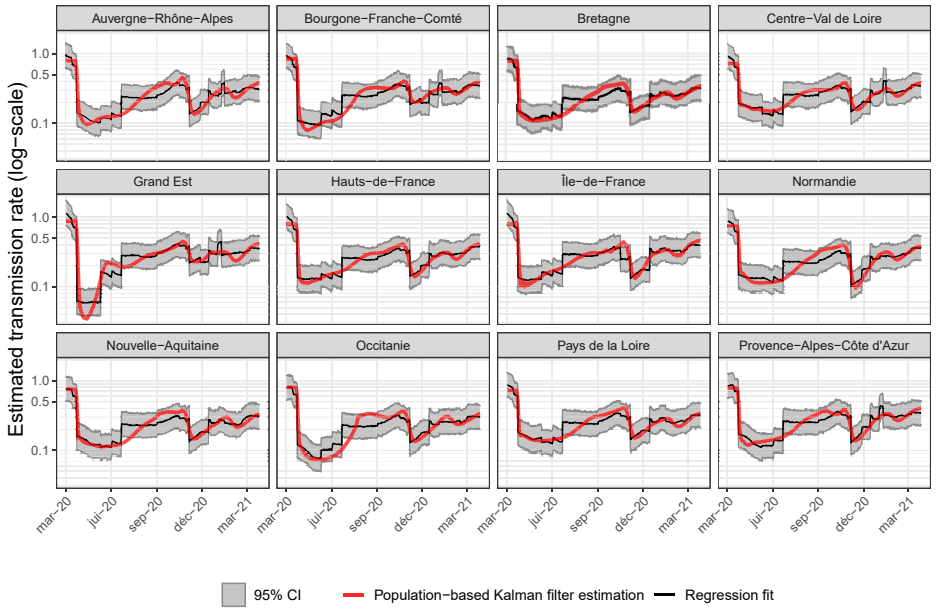


Figure 6: Regression fit of transmission rates from the population-based Kalman filter for each region with random effects on $\log(b_{init})$, b_{I1}, b_{I2} and $b_{curf6PM}$.

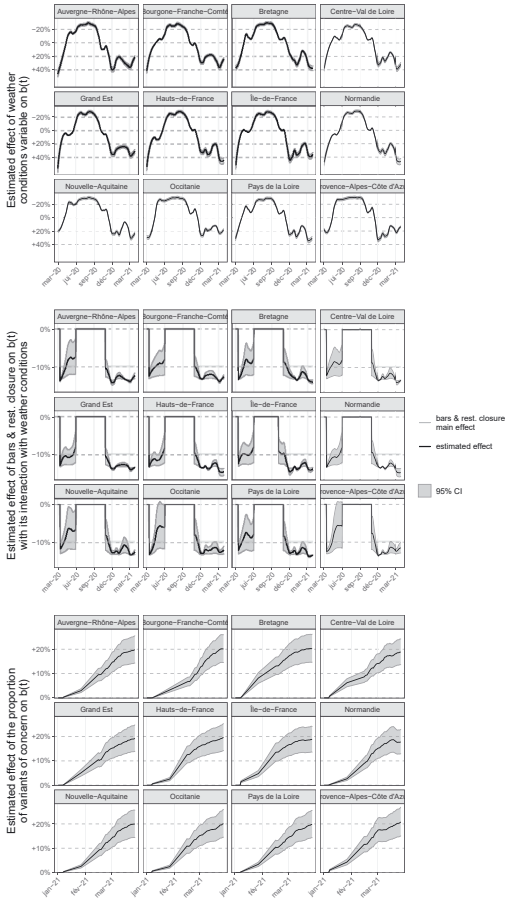


Figure 7: Top: estimated association between transmission rates and the seasonal weather conditions and their 95% confidence interval for the 12 regions of interest during the study period using the global average during the period as a reference for comparison. Middle: estimated association between transmission rates and bar and restaurant closures for the 12 regions of interest during the study period. In red, the main effect of −10%. In black and grey, the effect with the interaction with weather conditions and its 95% confidence band. Bottom: Estimated association between transmission rates and VoC occurrence and its 95% confidence interval from January 1, 2021.

seasonal weather conditions is statistically significant ($p = 0.037$), and complex to interpret. We show that while bar and restaurant closure always has a significant effect and reduces transmission, it is somewhat more effective in winter (11% [8%;14%] decrease) than in summer (8% [4%;11%] decrease). Figure 7 (middle) shows the estimated effect of bar and restaurant closures for the 12 regions of interest during the study period and their interaction with weather conditions. Coherently, the impact of weather conditions yields a stronger effect of bar and restaurant closures on reducing transmission in northern regions compared with southern regions (that can be interpreted as the closure being more effective in northern regions where weather conditions are also more favorable to transmission). In addition, a stronger effect is observed in spring 2021 due to weather more favorable to transmission compared to the previous year.

3.6 Effect of VoC

Regarding the effect of the three VoC, our results show that increasing the proportion from 0 to 100% would increase the transmission rate by an estimated 22% [15%; 28%]. Figure 7 (bottom) shows how the increased transmissibility varies over time due to the VoC fraction increasing to nearly 100%.

4 Discussion

We propose an innovative method to infer transmission rate over time from hospital data and estimate the association between transmission rates and multiple NPI, weather, and VoC. We show that all considered NPIs have a statistically significant and independent effect on transmission rates. In addition, we demonstrate a strong effect of weather conditions, decreasing transmission rate in summer and increasing it in winter, and we also observe increased transmissibility associated with VoC proportion increase.

Interpretation of our results on transmission rates is conditional on the mechanistic model presented in Figure 3 as well as in the parameter values set from the scientific literature and listed in Table 1. This model does not take into account the age structure of the population. However, we believe that using a population-based approach and relying on random effects takes into account several intrinsic characteristics – such as population density, age structure, transport habits, etc. – that influence the spread of disease through region-specific transmission rates. Another shortcoming of the *SEIRAH* model is that it does not account deterministically for travel between regions. However, since the state variables S, E, I, R, A, H are also corrected over time using the Kalman filters, our strategy can account for, for example, decreases or increases in the number of exposed individuals E in a region. This means that even if we do not directly model the interaction between regions, we are able to account for the most important changes that could be due to travel between regions. It could be further improved without changing the Kalman methodology by adding a modeling error operator applied specifically to this variable, provided we are able to quantify the degree of fluctuation in the population of each region due to travel. To do more, we could modify the model to include travel between regions, but this would require us to specify additional parameters that are difficult to calibrate without more data on travel between regions. Because the study period was short, vaccination was just starting over the end of the study period and immune escaping VoC were not circulating yet, waning immunity is also not taken into account. Regarding the study data, it is well known that there may be some misreporting in the SI-VIC database, especially when patients are transferred between hospitals or when the number of admissions is under reported during the summer months. However, one might expect these discrepancies to be less significant than for other data sources such as the daily number of new confirmed cases, where testing procedures (and availability) have changed significantly during the epidemic, or for primary care visits, where clinical diagnosis may be less accurate.

Interestingly, although the models were different and the data were not completely identical, our results were comparable in terms of attack rates to existing modeling work [63] and seroprevalence sampling studies [64–66] (for a comparison, see Appendix E). Overall, our estimates tend to be slightly higher, within a 5% margin. This is probably due to the strong assumption that there is no waning immunity in our model. Indeed, we assume that all individuals, once infected, have a sufficiently high antibody titers to systematically test positive in seroprevalence studies. Regarding the reproductive numbers, we estimate the average national basic reproductive number to be 3.10 [2.95; 3.26]. This differs from the initial values of R_{eff} in mid-March in all regions, which are estimated between 3.5 and 4, and is mainly due to winter weather conditions in mid-March in all regions. Our estimate is consistent with other estimates worldwide. In Liu et al. [67], the authors compare 12 studies that estimate the basic reproductive number for COVID-19 from China and overseas. Estimates ranged from 1.4 to 6.49, with a mean of 3.28. French studies also found similar estimates 3.18 [3.09; 3.24] in Di Domenico et al. [31] and 2.90 [2.81; 3.01] in Salje et al. [3]. These comparisons validate our estimation strategy, which has the major advantage of estimating transmission rates without making any assumptions about the shape of its dynamics over time. Moreover, the computation times are very reasonable (a few minutes for the full estimation on a regular laptop without code optimization).

Other variables could have been added, such as partial interventions (e.g., only in large cities in regions heavily affected by COVID − 19) − but these would have been inconsistent with our model defined at the regional level. Indeed, we need a match between hospital data-used to estimate transmission rates − and NPIs − used to estimate their associations with transmission rates − to combine the two steps of our study. To better illustrate, even if we have more information about bar closures in a city, we do not know exactly what proportion of that city's population is included in the hospital data for that city's region. And each region may have a different number of major cities, each of which has a different size in terms of population. So the impact of "partial bar closure" could vary from region to region, which is not consistent with the assumption of the regression.

Other potential variables were grouped under the NPI "barrier gestures" because they suffer from inconsistent definitions and are sometimes allocated simultaneously, leading to identifiability problems. Work from home is a good example. First, compliance with this measure (which can vary due to employee fatigue, organizational difficulties, and lack of legal requirements) is difficult to assess. Second, in terms of quantitative indicators, DARES has conducted surveys of companies with 10 or more employees and has shown that the average percentage of working from home is 30% [19%, 40%] and peaks during the two lockdowns [68]. However, this indicator takes into account work from home and paid leave, which can be very different in terms of risk behavior and overlap to some extent with school closure. Gathering restrictions is also a good example of how policies were constantly changing between private and public measures with the size of authorized gathering also varying. Finally, we believe that an individual's overall compliance with an entire package of "barrier gestures" is more likely to be stable over time than compliance with the individual components of that package. Mask wearing is a good example of this, as people were encouraged to keep the mask even though there was no longer an obligation, and probably did even more when at risk of infecting others. All in all, removing work from home, gathering restrictions, social distancing, mask wearing, and others from the regression model captures their association with transmission rates through the effect of lockdowns and barrier gestures in a more identifiable manner leading to more robust results.

We also make certain modeling assumptions. For some interventions, such as lockdowns, we considered a 7-day delay in implementation. This choice is not determined by the delay between infection and hospitalization, which is already accounted for by D_E and D_Q, but by the transmission rates obtained with the Kalman filter. This choice is supported by other studies, such as Dehning et al. [25], in which the delay can be as long as 15 days. However, we examined how this affects our regression fit. Not accounting for the 7-day lag significantly degrades the fits, see Model 4 of Appendix G in the Supporting Information for more details. Modeling the impact of partially opening schools during the period from May 11 to July 4, 2020, as 70% of the impact of a full closure might be considered an oversimplification. During this period after the first lockdown, schools reopened very gradually in three distinct phases, and enrollment increased even more progressively. By June 2nd, there were sharp differences between regions in opening and school attendance, with an average of only 30% of students under 12 attending school. The other levels of secondary schools ("collèges" and "lycées") reopened in early June and gradually enrolled more students until the vacations, which began in early July. To avoid risking identifiability problems, we choose not to differentiate by region or phase of reopening, and a ratio of 0.7 was used for the proportion of closure for all regions. We found modest but significant effect of school closure as of other studies [69]. Regarding the effect of curfews, we do not detect a statistically significant difference between 6 PM and 8 PM. This could be due to an identification problem, as the 8 PM curfew was in place for less than 3 weeks in many regions. An alternative explanation could be that both curfews affect global social gathering in the same way. For example, both prevent most private dinners and parties (or at least drastically reduce the number of guests).

We consider weather conditions relying on an aggregated indicator instead of exploring the impact of temperature, absolute humidity, and relative humidity separately. In Appendix G, simpler models are examined: all of them exhibits degraded fits. However, we remain extremely cautious in interpreting the estimated association of weather conditions with transmission rates and the mechanisms potentially involved. Finally, we consider an interaction between seasonal weather conditions and the closure of bars and restaurants, which is attributed to the use of patios and terraces (which have been expanded in many places since the

beginning of the pandemic). Of course, other interactions and more complex models can be considered, but this quickly leads to overfitting.

During our study period, the 20I/501Y.V1 (alpha) variant seems to have always been predominant (over 90% at any time) among VoC (the other two being beta and gamma) considered in mainland France [50]. Very different estimates of the increase in transmissibility for the 20I/501Y.V1 (alpha) variant have been proposed in the scientific literature, ranging from 29% to 90% [70–74], with the lower values being consistent with our own results. The higher estimates could be explained in part by the fact that new VoC appeared early in the winter in England and the United States. This could lead to confusion bias between weather conditions and VoC, both of which increase transmission, as mentioned by Campbell et al. [71]. We tested this assumption by removing weather conditions from our model and found an increased transmissibility of 43%, see Appendix G for VoC. Finally, we kept the weather conditions in our model because they greatly improved the regression fit.

We emphasize that our strategy does not allow us to estimate the direct effects of NPIs, weather, and new VoC on transmission rates, but is only suggestive of associations. Indeed, the only data available are observational, and the introduction of NPI is certainly not random, but clearly depends on the state of the epidemic. To estimate the direct effect, we would need to use either methods of causal inference with time-varying confounders or dynamic causality theory [75]. In the latter case, it is possible to use mechanistic models to estimate direct effects even in observational studies. However, transmission rates would have to be estimated directly in a one-step procedure using a parametric function that depends on these effects. This often relies on strong assumptions, notably the parametric shape of the transmission rate over time. But determining the shape of this dynamic is precisely the interest of this work using Kalman filters.

Overall, this work is one of the first attempts to retrospectively assess the associations between multiple NPIs and transmission rates over a one-year period of the COVID-19 epidemic. In addition to applying a novel methodology to a current and important application, this work could be extended to generate "what-if" scenarios and help determine appropriate NPI implementations for future waves of infection.

Acknowledgement: The authors thank the `opencovid-19` initiative for their contribution to the opening of the data used in this article. This work is supported in part by the Inria Mission COVID19, project GESTEPID. The authors sincerely thank Jane Heffernan for scientific discussions and thorough proofreading of the article. We also thank Linda Wittkop, Jane Heffernan, Quentin Clairon, Thomas Ferté, and Maria Pietro for constructive discussions about this work. Experiments presented in this paper were in part carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d'Aquitaine (see https://www.plafrim.fr).

Author contribution: AC, BPH, PM, MP and RT designed the study. LL and CV analyzed the data. AC, PM and MP implemented the software code. AC, BPH and MP interpreted the results. AC, BPH, LL, PM and MP wrote the manuscript.

Research funding: Part of the experiments presented in this paper were carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d'Aquitaine (see https://www.plafrim.fr).

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Appendix A: Supplementary figures for hospitalization, NPIs, variants of concern and vaccination data

Figure 8 represents the elsewhere published (SI-VIC database) and publicly available data on prevalence and incidence of hospitalization for COVID-19 in 12 non-insular French regions. Representation of NPIs in all regions is available in Figure 9. Finally, representations of the VoC proportion and the vaccination coverage ramping up (1st dose) in the population in each region over time are given in Figure 10.

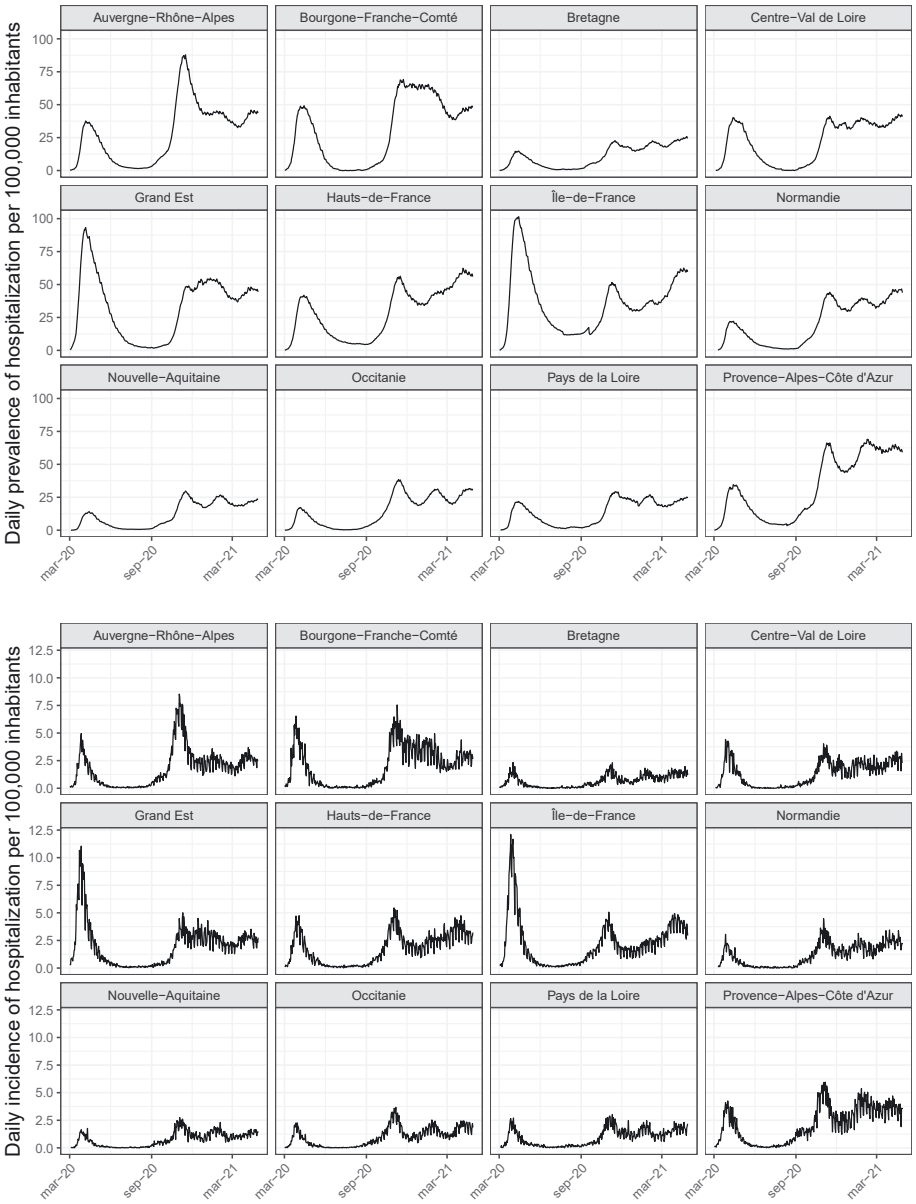


Figure 8: Top: total number of hospital bed occupied per 100,000 inhabitants ($100,000/N \times Y^{H}$). Bottom: daily number of new hospitalizations per 100,000 inhabitants $(100,000/N \times Y^{H_{\text{in}}})$.

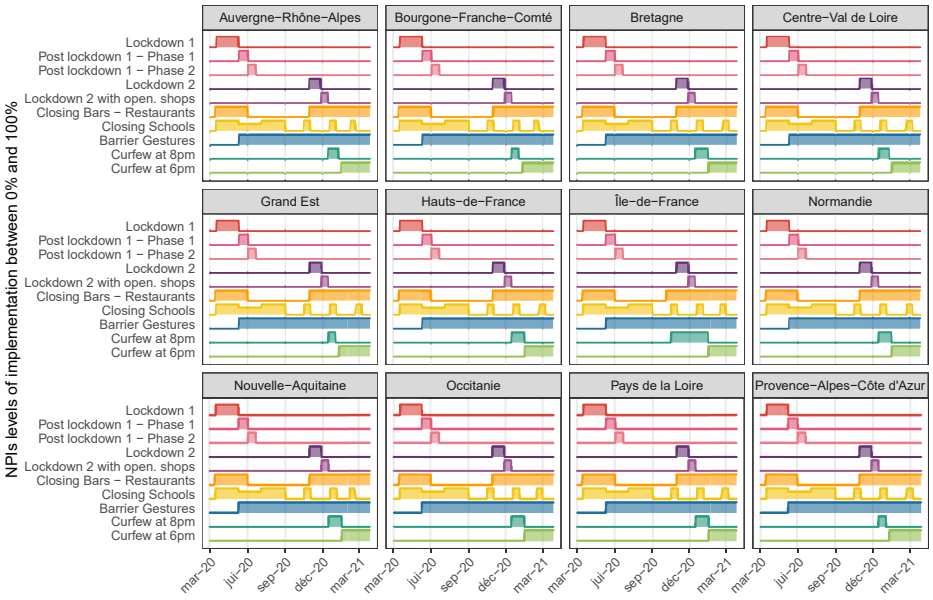


Figure 9: Implementation of NPIs in all regions. Differences are only impacting school closure and curfew by few days.

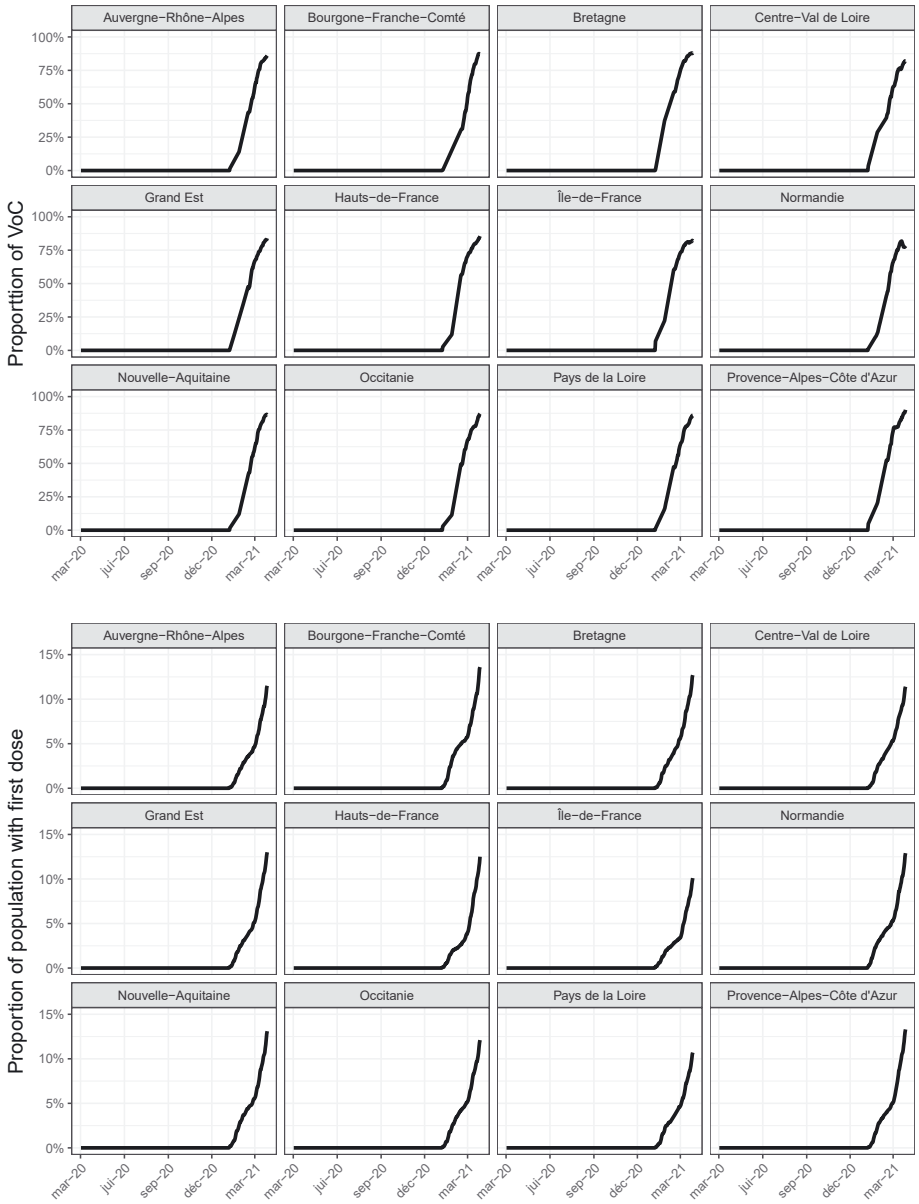


Figure 10: Top: Percentage of either UK, SA or BR VoC over time. Bottom: Percentage of people who have received a first dose of vaccine.

Appendix B: Estimation of the hospitalization period using the correlation between the total number of individual hospitalized daily and the daily incident number of hospitalization

The relation between the total number of individual hospitalized daily denoted by H and the daily incident number of hospitalization H_{in} is governed by

$$\frac{\mathrm{d}}{\mathrm{d}t}H = -\frac{H}{D_H} + H_{\mathrm{in}}, \tag{5}$$

with D_H the hospitalization period.

Using the data of daily incident number of hospitalization $\left(Y^{H_{\mathrm{in}}}\right)$ and the total number of individuals hospitalized daily (Y^H) over a period of about a year (391 days from March 2, 2020 to March 28, 2021), we estimate D_H in each region by using a mean squared estimation. The obtained values are given in Table 3. In the SEIRAH model, we fix D_H at the mean value.

Appendix C: Computation of the effective reproductive ratio

To compute the reproductive ratio R_{eff} of our SEIRAH model

$$\begin{cases} \dot{S} = -\underline{b}\left(1-\dfrac{V}{N}\right)\dfrac{S(I+\alpha A)}{N} \\ \dot{E} = \underline{b}\left(1-\dfrac{V}{N}\right)\dfrac{S(I+\alpha A)}{N} - \dfrac{E}{D_E} \\ \dot{I} = \dfrac{r_E}{D_E}E - \dfrac{1-r_I}{D_Q}I - \dfrac{r_I}{D_I}I, \\ \dot{R} = \dfrac{r_I I + A}{D_I} + \dfrac{H}{D_H} \\ \dot{A} = \dfrac{1-r_E}{D_E}E - \dfrac{A}{D_I} \\ \dot{H} = \dfrac{1-r_I}{D_Q}I - \dfrac{H}{D_H}, \end{cases} \tag{6}$$

we apply the *Next Generation Matrix* approach [76]. The principle consists in focusing on three categories: (i) latent E, (ii) ascertained infectious I and (iii) unascertained infectious A with the following dynamics

$$\begin{cases} \dfrac{\mathrm{d}E}{\mathrm{d}t} = \underline{b}\left(1-\dfrac{V}{N}\right)\dfrac{S(I+\alpha A)}{N} - \dfrac{E}{D_E} \\ \dfrac{\mathrm{d}I}{\mathrm{d}t} = \dfrac{r_E}{D_E}E - \dfrac{1-r_I}{D_Q}I - \dfrac{r_I}{D_I}I, \\ \dfrac{\mathrm{d}A}{\mathrm{d}t} = \dfrac{1-r_E}{D_E}E - \dfrac{A}{D_I} \end{cases}$$

Table 3: Estimation of D_H for the 12 regions: Île-de-France (IDF); Centre-Val de Loire (CVL); Bourgone-Franche-Comté (BFC); Normandie (Nor); Hauts-de-France (HDF); Grand Est (GE); Pays de la Loire (PL); Bretagne (Bret); Nouvelle-Aquitaine (NA); Occitanie (Occ); Auvergne-Rhône-Alpes (AURA); Provence-Alpes-Côte d'Azur (PACA).

	IDF	CVL	BFC	Norm.	HDF	GE	PL	Bret.	NA	Occ.	AURA	PACA	Nat. avg.
D_H (days)	18.3	19.5	18.0	20.6	18.6	17.7	16.7	19.6	18.1	17.4	17.0	18.1	18.3

Then, we build two matrices corresponding to: (1) V following the arrivals and departures from one other category and (2) F following the arrivals from another compartment exterior to the three categories. We have

$$V = \begin{pmatrix} \frac{1}{D_E} & 0 & 0 \\ -\frac{r_E}{D_E} & \frac{1-r_I}{D_Q} + \frac{r_I}{D_I} & 0 \\ -\frac{1-r_E}{D_E} & 0 & \frac{1}{D_I} \end{pmatrix} \quad \text{and} \quad F = \begin{pmatrix} 0 & \underline{b}\left(1-\frac{V}{N}\right)\frac{S}{N} & \alpha\underline{b}\left(1-\frac{V}{N}\right)\frac{S}{N} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It is then well known – see for instance Perasso et al. [77] for a proof – that

$$R_{\text{eff}} = \rho(\text{FV}^{-1}),$$

where $\rho(\text{FV}^{-1})$ is the spectral radius of the Next Generation Matrix FV^{-1}. Here, we have

$$\text{FV}^{-1} = \begin{pmatrix} \underline{b}\left(1-\frac{V}{N}\right)\frac{S}{N}\left(D_I\alpha(1-r_E) + \frac{D_I D_Q r_E}{(1-r_I)D_I + r_I D_Q}\right) & \underline{b}\left(1-\frac{V}{N}\right)\frac{S}{N}\frac{D_I D_Q}{(1-r_I)D_I + r_I D_Q} & \underline{b}\left(1-\frac{V}{N}\right)\frac{S}{N}D_I\alpha \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

with

$$V^{-1} = \begin{pmatrix} D_E & 0 & 0 \\ \frac{D_I D_Q r_E}{(1-r_I)D_I + D_Q r_I} & \frac{D_I D_Q}{(1-r_I)D_I + D_Q r_I} & 0 \\ (1-r_E)D_I & 0 & D_I \end{pmatrix}.$$

We therefore obtain

$$R_{\text{eff}}(t) = \underline{b}(t)\left(1-\frac{V(t)}{N}\right)\frac{S(t)}{N}\left(D_I\alpha(1-r_E) + \frac{D_I D_Q r_E}{(1-r_I)D_I + r_I D_Q}\right).$$

Appendix D: Initial transmission rate and attack rate estimated using our population-based Kalman filter

Table 4 shows the estimation of the initial values for the transmission rate at the regional level.

Table 4: Estimation of b_{init} for the 12 regions: Île-de-France (IDF); Centre-Val de Loire (CVL); Bourgone-Franche-Comté (BFC); Normandie (Nor); Hauts-de-France (HDF); Grand Est (GE); Pays de la Loire (PL); Bretagne (Bret); Nouvelle-Aquitaine (NA); Occitanie (Occ); Auvergne-Rhône-Alpes (AURA); Provence-Alpes-Côte d'Azur (PACA).

	IDF	CVL	BFC	Norm.	HDF	GE	PL	Bret.	NA	Occ.	AURA	PACA	Nat. avg.
$\boldsymbol{b}_{\text{init}}$	0.789	0.767	0.784	0.773	0.781	0.809	0.761	0.765	0.768	0.789	0.786	0.778	0.779

Appendix E: Comparison of obtained attack rates with other studies

Our attack rates are compared to (i) those obtained by Hoze et al. [63] (see Table 5), and to (ii) 3 seroprevalence studies [64–66] (see Table 6).

Table 5: Comparison of the estimated attack rates obtained in Hoze et al. [63] (first line) with our estimated attack rates (second line) at 3 dates for the metropolitan France.

	May 11, 2020	**October 31, 2020**	**January 15, 2021**
Hoze et al. [63]	5.7% [5.1%; 6.4%]	11% [9.7%; 12.4%]	14.9% [13.2%; 16.9%]
Proposed estimates	5.69% [5.61%; 5.77%]	12.78% [11.98%; 13.66%]	18.92% [16.76%; 21.43%]

Table 6: Comparison of the estimated attack rates obtained in 3 seroprevalence studies.

	May 2 – June 2, 2020	**May 4 – June 23, 2020**	**Oct. 5 – Oct. 11, 2020**
Auvergne-Rhône-Alpes*	4.8%	–	–
Auvergne-Rhône-Alpes**	4.48%	4.54%	7.15%
Bourgone-Franche-Comté*	1.5%	–	9.3%
Bourgone-Franche-Comté**	6.04%	6.14%	9.33%
Bretagne*	3.1%	–	–
Bretagne**	1.75%	1.79%	3.51%
Centre-Val de Loire*	2.1%	–	–
Centre-Val de Loire**	5.13%	5.26%	8.15%
Grand Est*	6.7%	9%	11.6%
Grand Est**	10.85%	11.36%	12.72%
Hauts-de-France*	2.9%	–	–
Hauts-de-France**	5.42%	5.64%	8.64%
Île-de-France*	9.2%	10%	14.8%
Île-de-France**	12.57%	13.06%	17.97%
Nouvelle-Aquitaine*	2.%	3.1%	–
Nouvelle-Aquitaine**	1.51%	1.52%	3.06%
Normandie*	1.9%	–	–
Normandie**	2.28%	2.31%	5.25%
Occitanie*	1.9%	–	–
Occitanie**	1.71%	1.74%	5.17%
Provence-Alpes-Côte d'Azur*	5.2%	–	–
Provence-Alpes-Côte d'Azur**	4.49%	4.63%	8.55%
Pays de la Loire*	3.4%	–	–
Pays de la Loire**	2.62%	2.69%	4.01%

*[64–66] (first line). **With our estimated attack rates (second line) averaged during the 3 corresponding date intervals.

Appendix F: Regression model

The model writes as follow for each region $i \in 1, \dots, 12$:

$$\begin{aligned}
\log(\underline{b}^i(t)) = {} & \alpha + \beta_{\text{Lock1}}\text{Lock1}(t) + \beta_{\text{Post-Lock1.1}}\text{Post} - \text{Lock1.1}^i(t) + \beta_{\text{Post-Lock1.2}}\text{Post} - \text{Lock1.2}^i(t) \\
& + \beta_{\text{Lock2}}\text{Lock2}^i(t) + \beta_{\text{Post-Lock2.1}}\text{Post} - \text{Lock2.1}^i(t) + \beta_{\text{ClosedSchools}}\text{ClosedSchools}^i(t) \\
& + \beta_{\text{ClosedBarsRestau}}\text{ClosedBarsRestau}^i(t) + \beta_{\text{BarrierGestures}}\text{BarrierGestures}^i(t) \\
& + \beta_{\text{Curf6PM}}\text{Curf6PM}^i(t) + \beta_{\text{Curf8PM}}\text{Curf8PM}^i(t) + \beta_{\text{VoC}}\text{VoC}^i(t) + \beta_{\text{W}}\text{W}^i(t) \\
& + \underbrace{\beta_{\text{Int}}\text{ClosedBarsRestau}^i(t) \times W^i(t)}_{\text{Interaction}} \\
& + \underbrace{\beta^i_{\text{Lock1}}\text{Lock1}^i(t) + \beta^i_{\text{Lock2}}\text{Lock2}^i(t) + \beta^i_{\text{Curf6PM}}\text{Curf6PM}^i(t)}_{\text{Random Slopes } \left(\beta^i_{\text{Lock1}}, \beta^i_{\text{Lock2}}, \beta^i_{\text{Lock1}}\right)^T \sim \mathcal{N}(0,\Sigma)} + \underbrace{u^i}_{\text{Randon Intercept } u_i \sim \mathcal{N}(0,\sigma_{\text{region}})}
\end{aligned} \tag{7}$$

Regression residuals, fixed and random effects are for the selected model given in Eq. (7) are given in Table 7. Covariance matrix is given in Table 8.

Table 7: Scaled residuals, random and fixed effects of regression model (7).

Scaled residuals					
	Min	1Q	Median	3Q	Max
	−3.9394	−0.4817	−0.0033	0.5609	10.5139
Random effects					
	Variance	Std. dev.	Corr		
σ_{region}	0.005159	0.07183			
σ_{Lock1}	0.106598	0.32649	−0.43		
σ_{Lock2}	0.018070	0.13443	0.35	−0.52	
$\sigma_{\text{Curf-6PM}}$	0.002827	0.05317	−0.29	0.12	−0.39
Fixed effects					
	Estimate	Std. error	Df	*t* value	Pr(>|*t*|)
A	−0.40919	0.02595	25.34463	−15.766	1.30e−14
Lock1	−1.52327	0.09631	11.76340	−15.817	2.75e−09
Post-Lock1.1	−0.77153	0.02037	4620.46129	−37.869	<2e−16
Post-Lock1.2	−0.65810	0.01427	4634.29686	−46.120	<2e−16
Lock2	−0.76626	0.04312	13.69861	−17.771	7.47e−11
Lock2.1	−0.71299	0.02169	4644.40638	−32.876	<2e−16
Closed schools	−0.07150	0.00866	4639.83981	−8.256	<2e−16
Closed bars & rest.	−0.10746	0.01492	4623.96873	−7.201	6.93e−13
Barrier gestures	−0.61300	0.01963	4652.90602	−31.229	<2e−16
Curf. 6 PM	−0.35386	0.02629	59.55020	−13.461	<2e−16
Curf. 8 PM	−0.32590	0.01951	4404.57885	−16.707	<2e−16
Variants	0.19505	0.02759	4605.07217	7.071	1.77e−12
Weather	−1.03117	0.04013	4654.84448	−25.696	<2e−16
Bar & rest.: weather	0.11877	0.05706	4501.30235	2.082	0.0374

Table 8: Covariance matrix of regression model (7).

	α	Lock1	Post-Lock1 – 1	Post-Lock1 – 2	Lock2	Lock2 – red.	Closed schools
α	6.7×10^{-4}	-9.9×10^{-4}	8.6×10^{-5}	-9.0×10^{-6}	3.7×10^{-4}	1.1×10^{-4}	-4.2×10^{-5}
Lock1	-9.9×10^{-4}	9.3×10^{-3}	1.2×10^{-4}	-1.6×10^{-6}	-1.9×10^{-3}	-1.9×10^{-5}	-1.5×10^{-5}
Post-Lock1.1	8.6×10^{-5}	1.2×10^{-4}	4.2×10^{-4}	6.0×10^{-6}	1.7×10^{-4}	1.2×10^{-4}	-1.9×10^{-5}
Post-Lock1.2	-9.0×10^{-6}	-1.6×10^{-6}	6.0×10^{-6}	2.0×10^{-4}	2.0×10^{-6}	2.1×10^{-6}	-4.2×10^{-6}
Lock2	3.7×10^{-4}	-1.9×10^{-3}	1.7×10^{-4}	2.0×10^{-6}	1.9×10^{-3}	2.1×10^{-4}	3.0×10^{-5}
Lock2.1	1.1×10^{-4}	-1.9×10^{-5}	1.2×10^{-4}	2.1×10^{-6}	2.1×10^{-4}	4.7×10^{-4}	2.6×10^{-5}
Closed schools	-4.2×10^{-5}	-1.5×10^{-5}	-1.9×10^{-5}	-4.2×10^{-6}	3.0×10^{-5}	2.6×10^{-5}	7.5×10^{-5}
Closed bars & rest.	-5.6×10^{-5}	-1.5×10^{-4}	-1.9×10^{-4}	1.5×10^{-5}	-1.5×10^{-4}	-1.2×10^{-4}	-1.7×10^{-5}
Barrier gestures	-2.7×10^{-4}	2.3×10^{-4}	-1.2×10^{-4}	-8.1×10^{-6}	-1.4×10^{-4}	-1.8×10^{-4}	2.9×10^{-5}
Curf. 6 PM	1.1×10^{-5}	2.1×10^{-4}	1.7×10^{-4}	2.7×10^{-6}	-1.7×10^{-5}	2.4×10^{-4}	2.1×10^{-5}
Curf. 8 PM	1.3×10^{-4}	-4.2×10^{-7}	1.4×10^{-4}	5.0×10^{-6}	1.9×10^{-4}	2.5×10^{-4}	-1.7×10^{-5}
VoC	3.7×10^{-5}	-5.7×10^{-5}	-5.3×10^{-5}	1.1×10^{-6}	1.0×10^{-5}	6.7×10^{-5}	-1.9×10^{-5}
Weather	2.5×10^{-4}	-2.7×10^{-4}	1.8×10^{-4}	-2.6×10^{-5}	1.4×10^{-4}	1.9×10^{-4}	-1.4×10^{-4}
Bar & rest.: weather	-8.5×10^{-5}	-1.1×10^{-4}	-5.5×10^{-4}	2.7×10^{-5}	-6.3×10^{-5}	2.3×10^{-4}	1.2×10^{-4}
	Closed bars & rest.	**Barrier gestures**	**Curf. 6 PM**	**Curf. 8 PM**	**Variants**	**Weather**	**Bar & rest.: weather**
A	-5.6×10^{-5}	-2.7×10^{-4}	1.1×10^{-5}	1.3×10^{-4}	3.7×10^{-5}	2.5×10^{-4}	-8.5×10^{-5}
Lock1	-1.5×10^{-4}	2.3×10^{-4}	2.1×10^{-4}	-4.2×10^{-7}	-5.7×10^{-5}	-2.7×10^{-4}	-1.1×10^{-4}
Post-Lock1.1	-1.9×10^{-4}	-1.2×10^{-4}	1.7×10^{-4}	1.4×10^{-4}	-5.3×10^{-5}	1.8×10^{-4}	-5.5×10^{-4}
Post-Lock1.2	1.5×10^{-5}	-8.1×10^{-6}	2.7×10^{-6}	5.0×10^{-6}	1.1×10^{-6}	-2.6×10^{-5}	2.7×10^{-5}
Lock2	-1.5×10^{-4}	-1.4×10^{-4}	-1.7×10^{-5}	1.9×10^{-4}	1.0×10^{-5}	1.4×10^{-4}	-6.3×10^{-5}
Lock2.1	-1.2×10^{-4}	-1.8×10^{-4}	2.4×10^{-4}	2.5×10^{-4}	6.7×10^{-5}	1.9×10^{-4}	2.3×10^{-4}
Closed schools	-1.7×10^{-5}	2.9×10^{-5}	2.1×10^{-5}	-1.7×10^{-5}	-1.9×10^{-5}	-1.4×10^{-4}	1.2×10^{-4}
Closed bars & rest.	2.2×10^{-4}	1.6×10^{-5}	-1.5×10^{-4}	-1.2×10^{-4}	3.3×10^{-5}	1.5×10^{-4}	5.1×10^{-5}
Barrier gestures	1.6×10^{-5}	3.9×10^{-4}	-1.7×10^{-4}	-2.0×10^{-4}	-4.7×10^{-5}	-5.2×10^{-4}	2.9×10^{-4}
Curf. 6 PM	-1.5×10^{-4}	-1.7×10^{-4}	6.9×10^{-4}	2.4×10^{-4}	-3.2×10^{-4}	1.8×10^{-4}	-8.1×10^{-6}
Curf. 8 PM	-1.2×10^{-4}	-2.0×10^{-4}	2.4×10^{-4}	3.8×10^{-4}	6.4×10^{-5}	2.8×10^{-4}	1.1×10^{-4}
VoC	3.3×10^{-5}	-4.7×10^{-5}	-3.2×10^{-4}	6.4×10^{-5}	7.6×10^{-4}	9.1×10^{-5}	2.9×10^{-4}
Weather	1.5×10^{-4}	-5.2×10^{-4}	1.8×10^{-4}	2.8×10^{-4}	9.1×10^{-5}	1.6×10^{-3}	-1.3×10^{-3}
Bar & rest.: weather	5.1×10^{-5}	2.9×10^{-4}	-8.1×10^{-6}	1.1×10^{-4}	2.9×10^{-4}	-1.3×10^{-3}	3.3×10^{-3}

Appendix G: Comparison with other regression models

In this part, we compare our regression model to other regression models. We start by considering a simple model neglecting the weather (Model 1) and then we consider a model integrating the weather variable but neglecting the interaction with the bars and restaurants (Model 2) and the selected model (Model 3). Model 4 corresponds to Model 3 but without considering the delay of 7 days after the lockdowns. Table 9 summarizes the results. Figure 11 shows the fits.

Using the first model (AIC $= -707$), we obtain a negative association between the closure of bars and restaurants and the transmission rates which is not realistic. This is due to the fact that the bars and restaurants were open during summer when the transmission rate was very low with a effective reproductive number inferior to 1 in all regions. That is why in a second model, we add the weather variable. The AIC of this model is larger superior to the first ones (AIC $= -1486$). The third model assumes that there is an interaction between the closures of bar and restaurants and the weather to take into account the use of terraces (which have been expanded in many places since the beginning of the pandemic). The AIC is similar to the second model (AIC $= -1485$). The AIC of Model 4 is very large (AIC $= 2224$) compared to other ones validating the delay of 7 days.

We found that the three VoCs (alpha, beta, and gamma) are $\sim$20% (Model 2 or 3) to $\sim$45% (Model 1) more transmissible than the historical lineage. The estimated value with Model 1 seems more realistic comparing to the literature and the end of the curve is better fitted. This is an important limitation of our work.

Table 9: Estimation of the associations between the transmission rates and the weather, the VoCs and the NPIs. Negative (resp. positive) values correspond to a decrease (resp. an increase) of the transmission rate.

NPI/variants/weather	Model 1	Model 2	Model 3	Model 4
Lockdown 1 (delay of 7 days)	−83% [−86%; −80%]	−78% [−82%; −74%]	−78% [−82%; −74%]	−65% [−71%; −58%]
Post lockdown 1 – phase 1	−54% [−56%; −53%]	−53% [−54%; −51%]	−54% [−56%; −52%]	−45% [−49%; −40%]
Post lockdown 1 – phase 2	−49% [−51%; −47%]	−48% [−50%; −47%]	−48% [−50%; −47%]	−48% [−50%; −46%]
Lockdown 2 (delay of 7 days)	−49% [−53%; −44%]	−53% [−57%; −49%]	−54% [−57%; −49%]	−41% [−46%; −35%]
Lockdown 2 with open. shops	−38% [−41%; −35%]	−51% [−53%; −49%]	−51% [−53%; −49%]	−54% [−57%; −51%]
Closing schools	−15% [−16%; −13%]	−7% [−9%; −6%]	−7% [−8%; −5%]	−3% [−6%; −1%]
Closing bars & restaurants	4% [1%; 7%]	−10% [−13%; −8%]	−10% [−13%; −8%]	−24% [−29%; −19%]
Barrier gestures	−63% [−64%; −61%]	−46% [−48%; −44%]	−46% [−48%; −44%]	−36% [−40%; −31%]
Curfew at 6 PM	−18% [−22%; −13%]	−30% [−33%; −26%]	−30% [−33%; −26%]	−28% [−34%; −23%]
Curfew at 8 PM	−5% [−8%; −2%]	−28% [−31%; −25%]	−28% [−31%; −25%]	−33% [−37%; −28%]
Proportion of variants	44% [36%; 52%]	20% [14%; 27%]	22% [15%; 28%]	6% [−2%; 15%]
Seasonal weather conditions	–	−63% [−65%; −60%]	−64% [−67%; −61%]	−73% [−76%; −69%]
Weather cond./closing bars & rest.	–	–	13% [1%; 26%]	−32% [−42%; −19%]
AIC	−707	−1486	−1485	2224

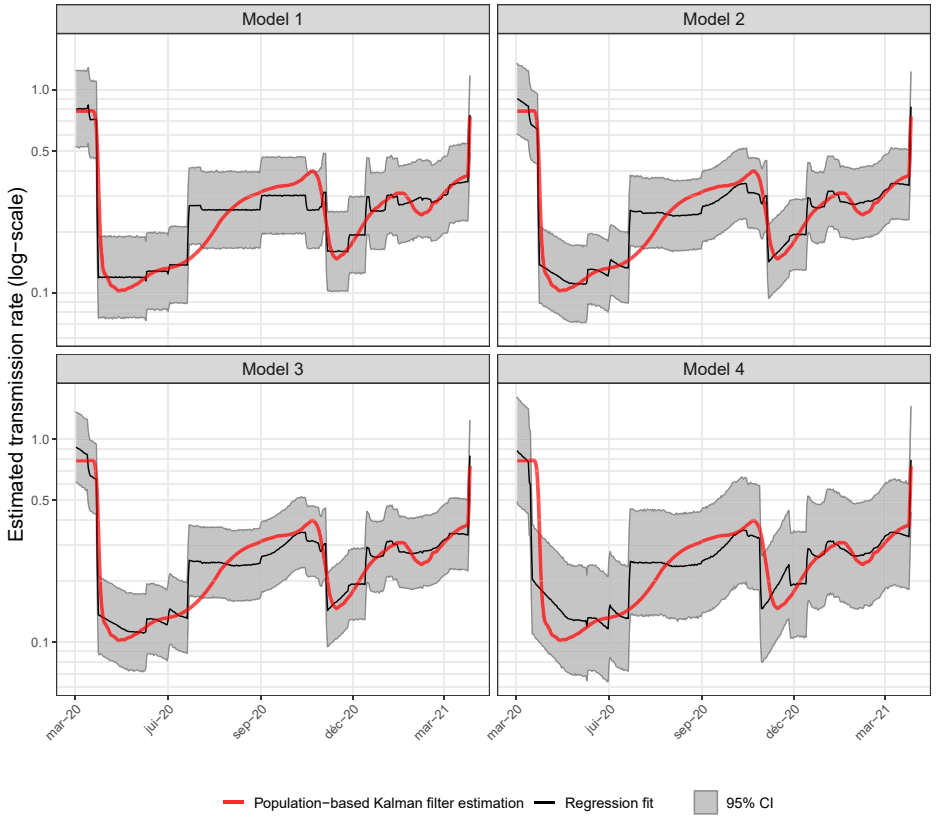


Figure 11: Results of regression models of the transmission rate obtained with the population based Kalman filter using Model 3 given in Eq. (7). Top-Left: Mean fit with linear model without considering weather effect (Model 1). Top-Right: Mean fit with linear model considering linear weather effect (Model 2). Bottom-Left: Mean fit with linear model with linear weather effect and with closed bar and restaurants and weather interaction (Model 3). Bottom-Right: Mean fit with linear model with linear weather effect and with closed bar and restaurants and weather interaction (Model 4) but without taking into account the delay of 7 days after the lockdowns. Random effects are considered on the intercept, the lockdowns and the curfew at 6 PM.

References

1. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72, 314 cases from the Chinese center for disease control and prevention. J Am Med Assoc 2020;323:1239–42.
2. Lapidus N, Paireau J, Levy-Bruhl D, de Lamballerie X, Severi G, Touvier M, et al. Do not neglect SARS-CoV-2 hospitalization and fatality risks in the middle-aged adult population. Infect Dis Now 2021;51:380–2.
3. Salje H, Tran Kiem C, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, et al. Estimating the burden of SARS-CoV-2 in France. Science 2020;369:208–11.
4. Angulo FJ, Finelli L, Swerdlow DL. Estimation of US SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths using seroprevalence surveys. JAMA Netw Open 2021;4:e2033706.
5. Docherty AB, Harrison EM, Green CA, Hardwick HE, Pius R, Norman L, et al. Features of 20 133 UK patients in hospital with COVID-19 using the ISARIC WHO clinical characterisation protocol: prospective observational cohort study. Br Med J 2020;369:m1985.

6. Jones TC, Biele G, Mühlemann B, Veith T, Schneider J, Beheim-Schwarzbach J, et al. Estimating infectiousness throughout SARS-CoV-2 infection course. Science 2021;373:eabi5273.
7. Wu P, Liu F, Chang Z, Lin Y, Ren M, Zheng C, et al. Assessing asymptomatic, pre-symptomatic and symptomatic transmission risk of SARS-CoV-2. Clin Infect Dis 2021;73:e1314—20.
8. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (Oxford COVID-19 government response tracker). Nat Human Behav 2021;5:529—38.
9. Davies NG, Kucharski AJ, Eggo RM, Gimma A, Edmunds WJ, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. Lancet Public Health 2020;5:e375—85.
10. Bubar KM, Reinholt K, Kissler SM, Lipsitch M, Cobey S, Grad YH, et al. Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. Science 2021;371:916—21.
11. Pei S, Kandula S, Shaman J. Differential effects of intervention timing on COVID-19 spread in the United States. Sci Adv 2020;6:eabd6370.
12. Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. Lancet Infect Dis 2021;21:193—202.
13. Banholzer N, van Weenen E, Lison A, Cenedese A, Seeliger A, Kratzwald B, et al. Estimating the effects of non-pharmaceutical interventions on the number of new infections with COVID-19 during the first epidemic wave. PLoS One 2021;16:e0252827.
14. Islam N, Sharp SJ, Chowell G, Shabnam S, Kawachi I, Lacey B, et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. Br Med J 2020;370:m2743.
15. Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 2020;584:262—7.
16. Askitas N, Tatsiramos K, Verheyden B. Estimating worldwide effects of non-pharmaceutical interventions on COVID-19 incidence and population mobility patterns using a multiple-event study. Sci Rep 2021;11:1—13.
17. Liu Y, Morgenstern C, Kelly J, Lowe R, CMMID COVID-19 Working Group, Jit M. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. BMC Med. 2021;19:1—12.
18. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. Nat Human Behav 2020;4:1303—12.
19. Flaxman S, Mishra S, Gandy A, Unwin HJT, Coupland H, Mellan TA, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature 2020;584:257—61.
20. Bo Y, Guo C, Lin C, Zeng Y, Li HB, Zhang Y, et al. Effectiveness of non-pharmaceutical interventions on COVID-19 transmission in 190 countries from 23 January to 13 April 2020. Int J Infect Dis 2021;102:247—53.
21. Sharma M, Mindermann S, Rogers-Smith C, Leech G, Snodin B, Ahuja J, et al. Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe. Nat Commun 2021;12:1—13.
22. Zhang XS, Vynnyvky E, Charlett A, De Angelis D, Chen Z, Liu W. Transmission dynamics and control measures of COVID-19 outbreak in China: a modelling study. Sci Rep 2021;11:2652.
23. Nader IW, Zeilinger EL, Jomar D, Zauchner C. Onset of effects of non-pharmaceutical interventions on COVID-19 infection rates in 176 countries. BMC Publ Health 2021;21:1—7.
24. Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 2020;368:493—7.
25. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science 2020;369:6500.
26. Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the effectiveness of government interventions against COVID-19. Science 2021;371:6531.
27. Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature 2020;585:410—3.
28. Ge Y, Zhang WB, Liu H, Ruktanonchai CW, Hu M, Wu X, et al. Impacts of worldwide individual non-pharmaceutical interventions on COVID-19 transmission across waves and space. Int J Appl Earth Obs Geoinf 2022;106:102649.
29. Bisanzio D, Reithinger R, Alqunaibet A, Almudarra S, Alsukait RF, Dong D, et al. Estimating the effect of non-pharmaceutical interventions to mitigate COVID-19 spread in Saudi Arabia. BMC Med 2022;20:1—14.
30. Ge Y, Zhang WB, Wu X, Ruktanonchai CW, Liu H, Wang J, et al. Untangling the changing impact of non-pharmaceutical interventions and vaccination on European Covid-19 trajectories. Nat Commun 2022;13:1—9.
31. Di Domenico L, Pullano G, Sabbatini CE, Boëlle PY, Colizza V. Impact of lockdown on COVID-19 epidemic in Île-de-France and possible exit strategies. BMC Med. 2020;18:1—13.
32. Collin A, Prague M, Moireau P. Estimation for dynamical systems using a population-based Kalman filter — applications in computational biology. MathS in Action 2022;11:213—42.
33. Simon D. Optimal state estimation: Kalman, H_∞, and nonlinear approaches. Oxford: Wiley—Blackwell; 2006.

34. Bensoussan A. Estimation and control of dynamical systems. In: Interdisciplinary applied mathematics. New York: Springer; 2018.
35. Lal R, Huang W, Li Z. An application of the ensemble Kalman filter in epidemiological modelling. PLoS One 2021;16:e0256227.
36. Arroyo-Marioli F, Bullano F, Kucinskas S, Rondón-Moreno C. Tracking R of COVID-19: a new real-time estimation using the Kalman filter. PLoS One 2021;16:e0244474.
37. Islam S, Hoque E, Amin MR. Integration of Kalman filter in the epidemiological model: a robust approach to predict COVID-19 outbreak in Bangladesh. Int J Mod Phys C 2021;32:2150108.
38. Hu B, Qiu W, Xu C, Wang J. Integration of a Kalman filter in the geographically weighted regression for modeling the transmission of hand, foot and mouth disease. BMC Publ. Health 2020;20:479.
39. Julier SJ, Uhlmann JK. Reduced sigma point filters for the propagation of means and covariances through nonlinear transformations. In: Proceedings of the 2002 American control conference; 2002. vol 2:887—92 pp.
40. French government's action summary website. Available from: https://www.gouvernement.fr/info-coronavirus/les-actions-du-gouvernement.
41. Zaitchik BF, Sweijd N, Shumake-Guillemot J, Morse A, Gordon C, Marty A, et al. A framework for research linking weather, climate and COVID-19. Nat Commun 2020;11:1—3.
42. Rodó X, San-José A, Kirchgatter K, López L. Changing climate and the COVID-19 pandemic: more than just heads or tails. Nat Med 2021;27:576—9.
43. Liu X, Huang J, Li C, Zhao Y, Wang D, Huang Z, et al. The role of seasonality in the spread of COVID-19 pandemic. Environ Res 2021;195:110874.
44. Islam N, Bukhari Q, Jameel Y, Shabnam S, Erzurumluoglu AM, Siddique MA, et al. COVID-19 and climatic factors: a global analysis. Environ Res 2021;193:110355.
45. Bukhari Q, Massaro JM, D'Agostino RB, Khan S. Effects of weather on coronavirus pandemic. Int J Environ Res Publ Health 2020;17:5399.
46. Roumagnac A, de Carvalho Filho E, Bertrand R, Banchereau AK, Lahache G. Étude de l'influence potentielle de l'humidité et de la température dans la propagation de la pandémie COVID-19. Med Catastr - Urgences Collect 2021;5:87—102.
47. World Health Organization. COVID-19 Weekly Epidemiological Update — 25 February 2021. Available from: https://apps.who.int/iris/bitstream/handle/10665/339859/nCoV-weekly-sitrep25Feb21-eng.pdf?sequence=1&isAllowed=y.
48. Santé Publique France. COVID-19 — Point épidémiologique hebdomadaire du 28 Janvier 2021. Available from: https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-28-janvier-2021.
49. Santé Publique France. COVID-19 — Point épidémiologique hebdomadaire du 11 Mars 2021. Available from: https://www.santepubliquefrance.fr/presse/2021/point-epidemiologique-covid-19-du-11-mars-2021.-circulation-virale-intense-tension-hospitaliere-accentuee-en-reanimation-variants-majoritaires.
50. Données de laboratoires pour le dépistage: Indicateurs sur les variants (SI-DEP). Available from: https://www.data.gouv.fr/en/datasets/.
51. Données relatives aux personnes vaccinées contre la COVID-19. Available from: https://www.data.gouv.fr/fr/datasets/donnees-relatives-aux-personnes-vaccinees-contre-la-covid-19-1/.
52. Pan A, Liu L, Wang C, Guo H, Hao X, Wang Q, et al. Association of public health interventions with the epidemiology of the COVID-19 outbreak wuhan, China. J Am Med Assoc 2020;323:1915—23.
53. Prague M, Wittkop L, Collin A, Clairon Q, Dutartre D, Moireau P, et al. Population modeling of early COVID-19 epidemic dynamics in French regions and estimation of the lockdown impact on infection rate. medRxiv 2020.
54. He J, Guo Y, Mao R, Zhang J. Proportion of asymptomatic coronavirus disease 2019: a systematic review and meta-analysis. J Med Virol 2021;93:820—30.
55. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science 2020;368:489—93.
56. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann Intern Med 2020;172:577—82.
57. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. The Lancet Microbe 2020;2:e13—22.
58. Delfraissy JF, Atlani Duault L, Benamouzig D, Bouadma L, Cauchemez S, Chauvin F, et al. Une deuxième vague entrainant une situation sanitaire critique. In: Note du conseil scientifique COVID-19; 2020.
59. Julier SJ, Uhlmann JK. A new extension of the Kalman filter to nonlinear systems. In: Signal processing, sensor fusion, and target recognition VI; 1997. vol 3068:182—93 pp.
60. Lavielle M. Mixed effects models for the population approach: models, tasks, methods and tools. Abingdon-on-Thames: CRC Press; 2014.

61. Bulfone TC, Malekinejad M, Rutherford GW, Razani N. Outdoor transmission of SARS-CoV-2 and other respiratory viruses: a systematic review. J Infect Dis 2021;223:550–61.
62. Adam D, Wu P, Peng W, Lau E, Tsang T, Cauchemez S, et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. Nature Medicine 2020;26:1714–9.
63. Hoze N, Paireau J, Lapidus N, Kiem CT, Salje H, Severi G, et al. Monitoring the proportion of the population infected by SARS-CoV-2 using age-stratified hospitalisation and serological data: a modelling study Lancet Public Health 2021;6:e408–15.
64. Warszawski J, Bajos N, Meyer L, de Lamballerie X, Seng R, Beaumont AL, et al. ER 1167: In May 2020, 4.5% of the population 2020 of metropolitan France had developed antibodies against SARS-CoV-2 — the first results of the EpiCov national survey. DREES - INSERM; 2020. Available from: https://drees.solidarites-sante.gouv.fr/sites/default/files/2020-10/er1167.pdf.
65. Carrat F, de Lamballerie X, Rahib D, Blanché H, Lapidus N, Artaud F, et al. Antibody status and cumulative incidence of SARS-CoV-2 infection among adults in three regions of France following the first lockdown and associated risk factors: a multicohort study. Int J Epidemiol 2021;50:1458–72.
66. Santé Publique France. COVID-19 — Point épidémiologique hebdomadaire du 31 Décembre 2020. Available from: https://www.santepubliquefrance.fr/maladies-et-traumatismes/maladies-et-infections-respiratoires/infection-a-coronavirus/documents/bulletin-national/covid-19-point-epidemiologique-du-31-decembre-2020.
67. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Trav Med 2020;27:taaa021.
68. Déchiffrer le monde du travail pour éclaire le débat public (DARES) - Principaux résultats des derniers mois - avril 2021. Available from: https://dares.travail-emploi.gouv.fr/publication/activite-et-conditions-demploi-de-la-main-doeuvre-pendant-la-crise-sanitaire-covid-19-1.
69. Stage H, Shingleton J, Ghosh S, Scarabel F, Pellis L, Finnie T. Shut and re-open: the role of schools in the spread of COVID-19 in Europe. Phil. Trans. Biol. Sci. 2020;376:0277.
70. Graham MS, Sudre CH, May A, Antonelli M, Murray B, Varsavsky T, et al. Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B.1.1.7: an ecological study. Lancet Public Health 2021;6:e335–45.
71. Campbell F, Archer B, Laurenson-Schafer H, Jinnai Y, Konings F, Batra N, et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. Euro Surveill 2021;26:2100509.
72. Washington NL, Gangavarapu K, Zeller M, Bolze A, Cirulli ET, Barrett KMS, et al. Emergence and rapid transmission of SARS-CoV-2 B. 1.1. 7 in the United States. Cell 2021;184:2587–94.
73. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature 2021;593:266–9.
74. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science 2021;372:eabg3055.
75. Prague M, Commenges D, Gran JM, Ledergerber B, Young J, Furrer H, et al. Dynamic models for estimating the effect of HAART on CD4 in observational studies: application to the aquitaine cohort and the Swiss HIV cohort study. Biometrics 2017;73:294–304.
76. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. J R Soc Interface 2005;2:281–93.
77. Perasso A. An introduction to the basic reproduction number in mathematical epidemiology. ESAIM: Proceedings and Surveys 2018;62:123–38.

Bibliography

[1] Odd O Aalen, Kjetil Røysland, Jon Michael Gran, and Bruno Ledergerber. Causality, mediation and time: a dynamic viewpoint. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 175(4):831–861, 2012.

[2] Emad Alamoudi, Yannik Schälte, Robert Müller, Jörn Starruß, Nils Bundgaard, Frederik Graw, Lutz Brusch, and Jan Hasenauer. Fitmulticell: simulating and parameterizing computational models of multi-scale and multi-cellular processes. *Bioinformatics*, 39(11):btad674, 2023.

[3] Kira Alhorn, Holger Dette, and Kirsten Schorning. Optimal designs for model averaging in non-nested models. *The indian journal of statistics*, 83(2):745–778, 2021.

[4] Samuel Alizon, Amy Hurford, Nicole Mideo, and Minus Van Baalen. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *Journal of evolutionary biology*, 22(2):245–259, 2009.

[5] Mathieu Andraud, Olivier Lejeune, Jammbe Z Musoro, Benson Ogunjimi, Philippe Beutels, and Niel Hens. Living on three time scales: the dynamics of plasma cell and antibody populations illustrated for hepatitis a virus. *PLoS computational biology*, 8(3):e1002418, 2012.

[6] Floriane Anstett-Collin, Lilianne Denis-Vidal, and Gilles Millérioux. A priori identifiability: An overview on definitions and approaches. *Annual Reviews in Control*, 50:139–149, 2020.

[7] Asier Antoranz, Carlos Mackintosh, Maria Ortiz, and Jon Pey. Elite: Expression deconvolution using linear optimization in bulk transcriptomics mixtures. *bioRxiv*, pages 2023–03, 2023.

[8] Elja Arjas and Jan Parner. Causal reasoning from longitudinal data. *Scandinavian Journal of Statistics*, 31(2):171–187, 2004.

[9] Mark Asch, Marc Bocquet, and Maëlle Nodet. *Data assimilation: methods, algorithms, and applications*. SIAM, 2016.

[10] Vincent Audigier and M Micemd Resche-Rigon. Multiple imputation by chained equations with multilevel data. *R package*, 2017.

[11] DJ Austin, NJ White, and RM Anderson. The dynamics of drug action on the within-host population growth of infectious agents: melding pharmacokinetics with pathogen population dynamics. *Journal of theoretical biology*, 194(3):313–339, 1998.

[12] Francisco Avila Cobos, Jo Vandesompele, Pieter Mestdagh, and Katleen De Preter. Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics*, 34(11):1969–1979, 2018.

[13] Géraldine Ayral, Jean-François Si Abdallah, Claude Magnard, and Jonathan Chauvin. A novel method based on unbiased correlations tests for covariate selection in nonlinear mixed effects models: The cossac approach. *CPT: pharmacometrics & systems pharmacology*, 10(4):318–329, 2021.

[14] Charline Bacchus-Souffan, Mark Fitch, Jori Symons, Mohamed Abdel-Mohsen, Daniel B Reeves, Rebecca Hoh, Mars Stone, Joseph Hiatt, Peggy Kim, Abha Chopra, et al. Relationship between cd4 t cell turnover, cellular differentiation and hiv persistence during art. *PLoS pathogens*, 17(1):e1009214, 2021.

[15] Harvey Thomas Banks, Marie Davidian, Shuhua Hu, Grace M Kepler, and Eric S Rosenberg. Modelling hiv immune response and validation with clinical data. *Journal of biological dynamics*, 2(4):357–385, 2008.

[16] Houreratou Barry, Gaudensia Mutua, Hannah Kibuuka, Zacchaeus Anywaine, Sodiomon B Sirima, Nicolas Meda, Omu Anzala, Serge Eholie, Christine Bétard, Laura Richert, et al. Safety and immunogenicity of 2-dose heterologous ad26. zebov, mva-bn-filo ebola vaccination in healthy and hiv-infected adults: A randomised, placebo-controlled phase ii clinical trial in africa. *PLoS medicine*, 18(10):e1003813, 2021.

[17] Giuseppina Bellu, Maria Pia Saccomani, Stefania Audoly, and Leontina D'Angiò. Daisy: A new software tool to test global identifiability of biological and physiological systems. *Computer methods and programs in biomedicine*, 88(1):52–61, 2007.

[18] Anne-Laure Boulesteix. Ten simple rules for reducing overoptimistic reporting in methodological computational research. *PLoS Computational Biology*, 11(4):e1004191, 2015.

[19] The BREATHER and Trial Group. Weekends-off efavirenz-based antiretroviral therapy in hiv-infected children, adolescents, and young adults (breather): a randomised, open-label, non-inferiority, phase 2/3 trial. *The lancet HIV*, 3(9):e421–e430, 2016.

[20] Karl Brendel, Emmanuelle Comets, Céline Laffont, Christian Laveille, and France Mentré. Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. *Pharmaceutical research*, 23:2036–2049, 2006.

[21] Simon Buatois, Sebastian Ueckert, Nicolas Frey, Sylvie Retout, and France Mentré. Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models. *The AAPS journal*, 20:1–9, 2018.

[22] Steven T Buckland, Kenneth P Burnham, and Nicole H Augustin. Model selection: an integral part of inference. *Biometrics*, pages 603–618, 1997.

[23] Laetitia Canini, Barbara Holzer, Sophie Morgan, Johanneke Dinie Hemmink, Becky Clark, sLoLa Dynamics Consortium, Mark EJ Woolhouse, Elma Tchilian, and Bryan Charleston. Timelines of infection and transmission dynamics of h1n1pdm09 in swine. *PLoS pathogens*, 16(7):e1008628, 2020.

[24] Louis Capitaine, Robin Genuer, and Rodolphe Thiébaut. Random forests for high-dimensional longitudinal data. *Statistical methods in medical research*, 30(1):166–184, 2021.

[25] Pierre Catoire, Robin Genuer, and Cécile Proust-Lima. Causal inference from observational data in emergency medicine research. *European Journal of Emergency Medicine*, 30(2):67–69, 2023.

[26] Amit Chandak, Debojyoti Dey, Bhaskar Mukhoty, and Purushottam Kar. Epidemiologically and socio-economically optimal policies via bayesian optimization. *Transactions of the Indian National Academy of Engineering*, 5(2):117–127, 2020.

[27] Shu-Hwa Chen, Wen-Yu Kuo, Sheng-Yao Su, Wei-Chun Chung, Jen-Ming Ho, Henry Horng-Shing Lu, and Chung-Yen Lin. A gene profiling deconvolution approach to estimating immune cell composition from complex tissues. *BMC bioinformatics*, 19:15–23, 2018.

[28] Lauren Childs, David W Dick, Zhilan Feng, Jane M Heffernan, Jing Li, and Gergely Röst. Modeling waning and boosting of covid-19 in canada with vaccination. *Epidemics*, 39:100583, 2022.

[29] Quentin Clairon. A regularization method for the parameter estimation problem in ordinary differential equations via discrete optimal control theory. *Journal of Statistical Planning and Inference*, 210:1–19, 2021.

[30] Jesse Clifton and Eric Laber. Q-learning: Theory and applications. *Annual Review of Statistics and Its Application*, 7:279–301, 2020.

[31] Stephen R Cole, Miguel A Hernán, Kathryn Anastos, Beth D Jamieson, and James M Robins. Determining the effect of highly active antiretroviral therapy on changes in human immunodeficiency virus type 1 rna viral load using a marginal structural left-censored mean model. *American journal of epidemiology*, 166(2):219–227, 2007.

[32] Annabelle Collin. Population-based estimation for pde system-applications in electroporation of tumor spheroids. *ESAIM: Control, Optimisation and Calculus of Variations*, 2024.

[33] Daniel Commenges and Anne Gégout-Petit. A general dynamical statistical model with causal interpretation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(3):719–736, 2009.

[34] Daniel Commenges and Anne Gégout-Petit. The stochastic system approach for estimating dynamic treatments effect. *Lifetime data analysis*, 21:561–578, 2015.

[35] Daniel Commenges, Cécile Proust-Lima, Cécilia Samieri, and Benoit Liquet. A universal approximate cross-validation criterion for regular risk functions. *The international journal of biostatistics*, 11(1):51–67, 2015.

[36] Kizzmekia S Corbett, Barbara Flynn, Kathryn E Foulds, Joseph R Francica, Seyhan Boyoglu-Barnum, Anne P Werner, Britta Flach, Sarah O'Connell, Kevin W Bock, Mahnaz Minai, et al. Evaluation of the mrna-1273 vaccine against sars-cov-2 in nonhuman primates. *New England Journal of Medicine*, 383(16):1544–1555, 2020.

[37] Daniel A Cruz and Melissa L Kemp. Hybrid computational modeling methods for systems biology. *Progress in Biomedical Engineering*, 4(1):012002, 2021.

[38] Charles Francis Curtiss and Joseph O Hirschfelder. Integration of stiff equations. *Proceedings of the national academy of sciences*, 38(3):235–243, 1952.

[39] Omid Dadras, Amir M Afsahi, Zahra Pashaei, Hengameh Mojdeganlou, Amirali Karimi, Pedram Habibi, Alireza Barzegary, Amirata Fakhfouri, Pegah Mirzapour, Nazanin Janfaza, et al. The relationship between covid-19 viral load and disease severity: a systematic review. *Immunity, inflammation and disease*, 10(3):e580, 2022.

[40] Gabriel Dagotto, Jingyou Yu, and Dan H Barouch. Approaches and challenges in sars-cov-2 vaccine development. *Cell host & microbe*, 28(3):364–370, 2020.

[41] Masoumeh Dashti, Kody JH Law, Andrew M Stuart, and Jochen Voss. Map estimators and their consistency in bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 2013.

[42] Courtney L Davis, Rezwanul Wahid, Franklin R Toapanta, Jakub K Simon, and Marcelo B Sztein. A clinically parameterized mathematical model of shigella immunity to inform vaccine design. *PloS one*, 13(1):e0189571, 2018.

[43] A Philip Dawid. Causal inference without counterfactuals. *Journal of the American statistical Association*, 95(450):407–424, 2000.

[44] Pierre de Truchis, Lambert Assoumou, Roland Landman, Dominique Mathez, Damien Le Dû, Jonathan Bellet, Karine Amat, Christine Katlama, Guillaume Gras, Olivier Bouchaud, et al. Four-days-a-week antiretroviral maintenance therapy in virologically controlled hiv-1-infected adults: the anrs 162-4d trial. *Journal of Antimicrobial Chemotherapy*, 73(3):738–747, 2018.

[45] Maud Delattre and Marc Lavielle. Coupling the saem algorithm and the extended kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Statistics and its interface*, 6(4):519–532, 2013.

[46] Maud Delattre, Marc Lavielle, and Marie-Anne Poursat. A note on bic in mixed-effects models. *Electronic Journal of Statistics*, 8:456–475, 2014.

[47] Solène Desmée, France Mentré, Christine Veyrat-Follet, Bernard Sébastien, and Jérémie Guedj. Using the saem algorithm for mechanistic joint models characterizing the relationship between nonlinear psa kinetics and survival in prostate cancer patients. *Biometrics*, 73(1):305–312, 2017.

[48] Karla Díaz-Ordaz, Michael G Kenward, Abie Cohen, Claire L Coleman, and Sandra Eldridge. Are missing data adequately handled in cluster randomised trials? a systematic review and guidelines. *Clinical Trials*, 11(5):590–600, 2014.

[49] Vanessa Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(1):245–264, 2008.

[50] Charlotte Dion, Simone Hermann, and Adeline Samson. Mixedsde: a r package to fit mixed stochastic differential equations. *The R Journal*, 11(1), 2019.

[51] Ruiwen Dong, Christian Goodbrake, Heather A Harrington, and Gleb Pogudin. Differential elimination for dynamical models via projections with applications to structural identifiability. *SIAM Journal on Applied Algebra and Geometry*, 7(1):194–235, 2023.

[52] Sophie Donnet and Adeline Samson. Parametric inference for mixed models defined by stochastic differential equations. *ESAIM: Probability and Statistics*, 12:196–218, 2008.

[53] Marie Doumic and Marc Hoffmann. Individual and population approaches for calibrating division rates in population dynamics: Application to the bacterial cell cycle. In *Modeling and Simulation for Collective Dynamics*, pages 1–81. World Scientific, 2023.

[54] Marie Doumic, Benoît Perthame, and Jorge P Zubelli. Numerical solution of an inverse problem in size-structured population dynamics. *Inverse Problems*, 25(4):045008, 2009.

[55] Julia Drylewicz, Jérémie Guedj, Daniel Commenges, and R Thiébaut. Modeling the dynamics of biomarkers during primary hiv infection taking into account the uncertainty of infection date. *The Annals of Applied Statistics*, pages 1847–1870, 2010.

[56] Wenjie Du, David Côté, and Yan Liu. Saits: Self-attention-based imputation for time series. *Expert Systems with Applications*, 219:119619, 2023.

[57] Cyrielle Dumont, Giulia Lestini, Hervé Le Nagard, France Mentré, Emmanuelle Comets, Thu Thuy Nguyen, et al. Pfim 4.0, an extended r program for design evaluation and optimization in nonlinear mixed-effect models. *Computer methods and programs in biomedicine*, 156:217–229, 2018.

[58] Kristen A Earle, Donna M Ambrosino, Andrew Fiore-Gartland, David Goldblatt, Peter B Gilbert, George R Siber, Peter Dull, and Stanley A Plotkin. Evidence for antibody as a protective correlate for covid-19 vaccines. *Vaccine*, 39(32):4423–4428, 2021.

[59] Baptiste Elie, Bénédicte Roquebert, Mircea T Sofonea, Sabine Trombert-Paolantoni, Vincent Foulongne, Jérémie Guedj, Stpéhanie Haim-Boukobza, and Samuel Alizon. Variant-specific sars-cov-2 within-host kinetics. *Journal of Medical Virology*, 94(8):3625–3633, 2022.

[60] Benjamin Engelhardt, Maik Kschischo, and Holger Fröhlich. A bayesian approach to estimating hidden variables as well as missing and wrong molecular interactions in ordinary differential equation-based mathematical models. *Journal of The Royal Society Interface*, 14(131):20170332, 2017.

[61] Heinz W Engl, Christoph Flamm, Philipp Kügler, James Lu, Stefan Müller, and Peter Schuster. Inverse problems in systems biology. *Inverse Problems*, 25(12):123014, 2009.

[62] Nir Eyal and Marc Lipsitch. How to test severe acute respiratory syndrome coronavirus 2 vaccines ethically even after one is available. *Clinical Infectious Diseases*, 73(12):2332–2334, 2021.

[63] Lucie Fayette, Romain Leroux, France Mentré, and Jérémy Seurat. Robust and adaptive two-stage designs in nonlinear mixed effect models. *The AAPS Journal*, 25(4):71, 2023.

[64] Shuo Feng, Daniel J Phillips, Thomas White, Homesh Sayal, Parvinder K Aley, Sagida Bibi, Christina Dold, Michelle Fuskova, Sarah C Gilbert, Ian Hirsch, et al. Correlates of protection against symptomatic and asymptomatic sars-cov-2 infection. *Nature medicine*, 27(11):2032–2040, 2021.

[65] Thomas Ferté, Vianney Jouhet, Romain Griffier, Boris P Hejblum, and Rodolphe Thiébaut. The benefit of augmenting open data with clinical data-warehouse ehr for forecasting sars-cov-2 hospitalizations in bordeaux area, france. *JAMIA open*, 5(4):ooac086, 2022.

[66] Mallorie H Fiero, Shuang Huang, Eyal Oren, and Melanie L Bell. Statistical analysis and handling of missing data in cluster randomized trials: a systematic review. *Trials*, 17:1–10, 2016.

[67] Thomas R Fleming and John H Powers. Biomarkers and surrogate endpoints in clinical trials. *Statistics in medicine*, 31(25):2973–2984, 2012.

[68] Christophe Fraser, Joanne E Tomassini, Liwen Xi, Greg Golm, Michael Watson, Anna R Giuliano, Eliav Barr, and Kevin A Ault. Modeling the long-term antibody response of a human papillomavirus (hpv) virus-like particle (vlp) type 16 prophylactic vaccine. *Vaccine*, 25(21):4324–4333, 2007.

[69] Sebastian Funk, Shweta Bansal, Chris T Bauch, Ken TD Eames, W John Edmunds, Alison P Galvani, and Petra Klepac. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics*, 10:21–25, 2015.

[70] Irene Garcia-Fogeda, Hajar Besbassi, Ynke Larivière, Benson Ogunjimi, Steven Abrams, and Niel Hens. Within-host modeling to measure dynamics of antibody responses after natural infection or vaccination: A systematic review. *Vaccine*, 2023.

[71] Robin Genuer, Jean-Michel Poggi, Robin Genuer, and Jean-Michel Poggi. *Random forests*. Springer, 2020.

[72] Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with lstm. *Neural computation*, 12(10):2451–2471, 2000.

[73] Mohammad Ghavamzadeh, Shie Mannor, Joelle Pineau, Aviv Tamar, et al. Bayesian reinforcement learning: A survey. *Foundations and Trends® in Machine Learning*, 8(5-6):359–483, 2015.

[74] Subhashis Ghosal. Normal approximation to the posterior distribution for generalized linear models with many covariates. *Mathematical Methods of Statistics*, 6(3):332–348, 1997.

[75] Peter B Gilbert, David C Montefiori, Adrian B McDermott, Youyi Fong, David Benkeser, Weiping Deng, Honghong Zhou, Christopher R Houchens, Karen Martins, Lakshmi Jayashankar, et al. Immune correlates analysis of the mrna-1273 covid-19 vaccine efficacy clinical trial. *Science*, 375(6576):43–50, 2022.

[76] Antonio Gonçalves, France Mentré, Annabelle Lemenuel-Diot, and Jérémie Guedj. Model averaging in viral dynamic models. *The AAPS Journal*, 22:1–11, 2020.

[77] Ashish Goyal, Daniel B Reeves, E Fabian Cardozo-Ojeda, Joshua T Schiffer, and Bryan T Mayer. Viral load and contact heterogeneity predict sars-cov-2 transmission and super-spreading events. *Elife*, 10:e63537, 2021.

[78] Frederik Graw and Alan S Perelson. Modeling viral spread. *Annual review of virology*, 3:555–572, 2016.

[79] Qian Guan, Brian J Reich, Eric B Laber, and Dipankar Bandyopadhyay. Bayesian nonparametric policy search with application to periodontal recall intervals. *Journal of the American Statistical Association*, 115(531):1066–1078, 2020.

[80] Jérémie Guedj, Rodolphe Thiébaut, and Daniel Commenges. Practical identifiability of hiv dynamics models. *Bulletin of mathematical biology*, 69:2493–2513, 2007.

[81] Trevor Hastie and Robert Tibshirani. Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398):371–386, 1987.

[82] Ana Maria Henao-Restrepo, Anton Camacho, Ira M Longini, Conall H Watson, W John Edmunds, Matthias Egger, Miles W Carroll, Natalie E Dean, Ibrahima Diatta, Moussa Doumbia, et al. Efficacy and effectiveness of an rvsv-vectored vaccine in preventing ebola virus disease: final results from the guinea ring vaccination, open-label, cluster-randomised trial (ebola ça suffit!). *The Lancet*, 389(10068):505–518, 2017.

[83] Miguel A Hernán, Babette A Brumback, and James M Robins. Estimating the causal effect of zidovudine on cd4 count with a marginal structural model for repeated measures. *Statistics in medicine*, 21(12):1689–1709, 2002.

[84] Alison L Hill, Daniel IS Rosenbloom, Feng Fu, Martin A Nowak, and Robert F Siliciano. Predicting the outcomes of treatment to eradicate the latent reservoir for hiv-1. *Proceedings of the National Academy of Sciences*, 111(37):13475–13480, 2014.

[85] David D Ho, Avidan U Neumann, Alan S Perelson, Wen Chen, John M Leonard, and Martin Markowitz. Rapid turnover of plasma virions and cd4 lymphocytes in hiv-1 infection. *Nature*, 373(6510):123–126, 1995.

[86] Rune Hoff, Jon Michael Gran, and Daniel Farewell. Farewell's linear increments model for missing data: The flim package. *R Journal*, 6(2), 2014.

[87] Hoon Hong, Alexey Ovchinnikov, Gleb Pogudin, and Chee Yap. Global identifiability of differential models. *Communications on Pure and Applied Mathematics*, 73(9):1831–1879, 2020.

[88] Giles Hooker and Stephen P Ellner. Goodness of fit in a nonlinear dynamics: misspecified rates or misspecified states? *THE ANNALS of APPLIED STATISTICS*, pages 754–776, 2015.

[89] Yangxin Huang and Tao Lu. Modeling long-term longitudinal hiv dynamics with application to an aids clinical study. *Annals of Applied Statistics*, 2(4):1384–1408, 2008.

[90] Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

[91] Md Hamidul Huque, John B Carlin, Julie A Simpson, and Katherine J Lee. A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC medical research methodology*, 18:1–16, 2018.

[92] Louise Madeleine Ince, Coline Barnoud, Lydia Kay Lutes, Robert Pick, Chen Wang, Flore Sinturel, Chien-Sin Chen, Alba de Juan, Jasmin Weber, Stephan J Holtkamp, et al. Influence of circadian clocks on adaptive immunity and vaccination responses. *Nature communications*, 14(1):476, 2023.

[93] David Ishola, Daniela Manno, Muhammed O Afolabi, Babajide Keshinro, Viki Bockstal, Baimba Rogers, Kwabena Owusu-Kyei, Alimamy Serry-Bangura, Ibrahim Swaray, Brett Lowe, et al. Safety and long-term immunogenicity of the two-dose heterologous ad26. zebov and mva-bn-filo ebola vaccine regimen in adults in sierra leone: A combined open-label, non-randomised stage 1, and a randomised, double-blind, controlled stage 2 trial. *The Lancet Infectious Diseases*, 22(1):97–109, 2022.

[94] Kazuyuki Iwata, Kohkichi Kawasaki, and Nanako Shigesada. A dynamical model for the growth and size distribution of multiple metastatic tumors. *Journal of theoretical biology*, 203(2):177–186, 2000.

[95] David LI Janzén, Mats Jirstrand, Michael J Chappell, and Neil D Evans. Three novel approaches to structural identifiability analysis in mixed-effects models. *Computer methods and programs in biomedicine*, 171:141–152, 2019.

[96] Pengfei Jin, Jingxin Li, Hongxing Pan, Yanfei Wu, and Fengcai Zhu. Immunological surrogate endpoints of covid-2019 vaccines: the evidence we have versus the evidence we need. *Signal transduction and targeted therapy*, 6(1):48, 2021.

[97] Marshall M Joffe and Tom Greene. Related causal frameworks for surrogate outcomes. *Biometrics*, 65(2):530–538, 2009.

[98] E Niclas Jonsson and Mats O Karlsson. Automated covariate model building within nonmem. *Pharmaceutical research*, 15:1463–1468, 1998.

[99] D Joubert, JD Stigter, and J Molenaar. An efficient procedure to assist in the re-parametrization of structurally unidentifiable models. *Mathematical biosciences*, 323:108328, 2020.

[100] Boris Julg, Lynda Dee, Jintanat Ananworanich, Dan H Barouch, Katharine Bar, Marina Caskey, Donn J Colby, Liza Dawson, Krista L Dong, Karine Dubé, et al. Recommendations for analytical antiretroviral treatment interruptions in hiv research trials—report of a consensus meeting. *The lancet HIV*, 6(4):e259–e268, 2019.

[101] Boris Julg, Kathryn E Stephenson, Kshitij Wagh, Sabrina C Tan, Rebecca Zash, Stephen Walsh, Jessica Ansel, Diane Kanjilal, Joseph Nkolola, Victoria EK Walker-Sperling, et al. Safety and antiviral activity of triple combination broadly neutralizing monoclonal antibody therapy against hiv-1: a phase 1 clinical trial. *Nature medicine*, 28(6):1288–1296, 2022.

[102] Simon J Julier and Jeffrey K Uhlmann. New extension of the kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI*, volume 3068, pages 182–193. Spie, 1997.

[103] Christel Kamp, Mathieu Moslonka-Lefebvre, and Samuel Alizon. Epidemic spread on weighted networks. *PLoS computational biology*, 9(12):e1003352, 2013.

[104] Melissa Kapulu, Lucinda Manda-Taylor, Shobana Balasingam, Gary Means, Mikal Ayiro Malungu, Philip Bejon, Primus Che Chi, Christopher Chiu, E Chandler Church, Rodrigo Correa-Oliveira, et al. Fourth controlled human infection model (chim) meeting–chims in endemic countries, may 22–23, 2023. *Biologicals*, page 101747, 2024.

[105] David S Khoury, Deborah Cromer, Arnold Reynaldi, Timothy E Schlub, Adam K Wheatley, Jennifer A Juno, Kanta Subbarao, Stephen J Kent, James A Triccas, and Miles P Davenport. Neutralizing antibody levels are highly predictive of immune protection from symptomatic sars-cov-2 infection. *Nature medicine*, 27(7):1205–1211, 2021.

[106] Ben Killingley, Alex J Mann, Mariya Kalinova, Alison Boyers, Niluka Goonawardane, Jie Zhou, Kate Lindsell, Samanjit S Hare, Jonathan Brown, Rebecca Frise, et al. Safety, tolerability and viral kinetics during sars-cov-2 human challenge in young adults. *Nature Medicine*, 28(5):1031–1041, 2022.

[107] Denise Kirschner, Suzanne Lenhart, and Steve Serbin. Optimal control of the chemotherapy of hiv. *Journal of mathematical biology*, 35:775–792, 1997.

[108] Søren Klim, Stig Bousgaard Mortensen, Niels Rode Kristensen, Rune Viig Overgaard, and Henrik Madsen. Population stochastic modelling (psm)—an r package for mixed-effects models based on stochastic differential equations. *Computer methods and programs in biomedicine*, 94(3):279–289, 2009.

[109] Cyrille Kone, Emilie Kaufmann, and Laura Richert. Adaptive algorithms for relaxed pareto set identification. *Advances in Neural Information Processing Systems*, 36:35190–35201, 2023.

[110] Kenneth G Kowalski and Matthew M Hutmacher. Efficient screening of covariates in population models using wald's approximation to the likelihood ratio test. *Journal of pharmacokinetics and pharmacodynamics*, 28:253–275, 2001.

[111] Estelle Kuhn and Marc Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational statistics & data analysis*, 49(4):1020–1038, 2005.

[112] Massimiliano Lanzafame, Emanuela Lattuada, Dora Luise, Andrea Delama, Daniela Fait, and Sandro Vento. Short cycle, intermittent therapy: A valuable option in selected, virologically suppressed people living with hiv. *AIDS Research and Human Retroviruses*, 40(2):69–72, 2024.

[113] Reinhard Laubenbacher, Anna Niarakis, Tomáš Helikar, Gary An, Bruce Shapiro, Rahuman S Malik-Sheriff, TJ Sego, Adam Knapp, Paul Macklin, and James A Glazier. Building digital twins of the human immune system: toward a roadmap. *NPJ digital medicine*, 5(1):64, 2022.

[114] Marc Lavielle and Leon Aarons. What do we mean by identifiability in mixed effects models? *Journal of pharmacokinetics and pharmacodynamics*, 43:111–122, 2016.

[115] Marc Lavielle and Cyprien Mbogning. An improved saem algorithm for maximum likelihood estimation in mixtures of non linear mixed effects models. *Statistics and Computing*, 24:693–707, 2014.

[116] Marc Lavielle and France Mentré. Estimation of population pharmacokinetic parameters of saquinavir in hiv patients with the monolix software. *Journal of pharmacokinetics and pharmacodynamics*, 34:229–249, 2007.

[117] Marc Lavielle, Adeline Samson, Ana Karina Fermin, and France Mentré. Maximum likelihood estimation of long-term hiv dynamic models and antiviral response. *Biometrics*, 67(1):250–259, 2011.

[118] Francois-Xavier Lescure, Lila Bouadma, Duc Nguyen, Marion Parisey, Paul-Henri Wicky, Sylvie Behillil, Alexandre Gaymard, Maude Bouscambert-Duchamp, Flora Donati, Quentin Le Hingrat, et al. Clinical and virological data of the first cases of covid-19 in europe: a case series. *The Lancet Infectious Diseases*, 20(6):697–706, 2020.

[119] Zhen Li, Jie Chen, Eric Laber, Fang Liu, and Richard Baumgartner. Optimal treatment regimes: a review and empirical comparison. *International Statistical Review*, 2023.

[120] Pieter JK Libin, Arno Moonens, Timothy Verstraeten, Fabian Perez-Sanjines, Niel Hens, Philippe Lemey, and Ann Nowé. Deep reinforcement learning for large-scale epidemic control. *ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part V*, pages 155–170, 2021.

[121] Thomas S Ligon, Fabian Fröhlich, Oana T Chiş, Julio R Banga, Eva Balsa-Canto, and Jan Hasenauer. Genssi 2.0: multi-experiment structural identifiability analysis of sbml models. *Bioinformatics*, 34(8):1421–1423, 2018.

[122] Lijing Lin, Matthew Sperrin, David A Jenkins, Glen P Martin, and Niels Peek. A scoping review of causal methods enabling predictions under hypothetical interventions. *Diagnostic and prognostic research*, 5:1–16, 2021.

[123] Dora Luise, Emanuela Lattuada, Sebastiano Rizzardo, Stefano Nicolè, Lorenza Lambertenghi, Ilaria Coledan, Silvia Gambino, Rossella Gottardo, Massimiliano Lanzafame, and Sandro Vento. Short-cycle therapy in hiv-infected adults: rilpivirine combination 4 days on/3 days off therapy. *Journal of Antimicrobial Chemotherapy*, 77(3):747–752, 2022.

[124] Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer science review*, 3(3):127–149, 2009.

[125] David J Lunn, Andrew Thomas, Nicky Best, and David Spiegelhalter. Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10:325–337, 2000.

[126] Vincent Madelain, France Mentré, Sylvain Baize, Xavier Anglaret, Cédric Laouénan, Lisa Oestereich, Thi Huyen Tram Nguyen, Denis Malvy, Géraldine Piorkowski, Frederik Graw, et al. Modeling favipiravir antiviral efficacy against emerging viruses: from animal studies to clinical trials. *CPT: pharmacometrics & systems pharmacology*, 9(5):258–271, 2020.

[127] Tim Maiwald, Helge Hass, Bernhard Steiert, Joep Vanlier, Raphael Engesser, Andreas Raue, Friederike Kipkeew, Hans H Bock, Daniel Kaschek, Clemens Kreutz, et al. Driving the model to its limit: profile likelihood based model reduction. *PloS one*, 11(9):e0162366, 2016.

[128] Jaap W Mandema, Davide Verotta, and Lewis B Sheiner. Building population pharmacokinetic-pharmacodynamic models. i. models for covariate effects. *Journal of pharmacokinetics and biopharmaceutics*, 20(5):511–528, 1992.

[129] Aurélien Marc, Marion Kerioui, François Blanquart, Julie Bertrand, Oriol Mitja, Marc Corbacho-Monné, Michael Marks, and Jeremie Guedj. Quantifying the relationship between sars-cov-2 viral load and infectiousness. *Elife*, 10:e69302, 2021.

[130] Gemma Massonis, Julio R Banga, and Alejandro F Villaverde. Structural identifiability and observability of compartmental models of the covid-19 pandemic. *Annual reviews in control*, 51:441–459, 2021.

[131] Laura Matrajt, Julia Eaton, Tiffany Leung, Dobromir Dimitrov, Joshua T Schiffer, David A Swan, and Holly Janes. Optimizing vaccine allocation for covid-19 vaccines shows the potential role of single-dose vaccination. *Nature communications*, 12(1):3449, 2021.

[132] Laura Matrajt, Holly Janes, Joshua T Schiffer, and Dobromir Dimitrov. Quantifying the impact of lifting community nonpharmaceutical interventions for covid-19 during vaccination rollout in the united states. In *Open forum infectious diseases*, volume 8(7), page ofab341. Oxford University Press US, 2021.

[133] Katherine McMahan, Jingyou Yu, Noe B Mercado, Carolin Loos, Lisa H Tostanoski, Abishek Chandrashekar, Jinyan Liu, Lauren Peter, Caroline Atyeo, Alex Zhu, et al. Correlates of protection against sars-cov-2 in rhesus macaques. *Nature*, 590(7847):630–634, 2021.

[134] France Mentre, Alain Mallet, and Doha Baccar. Optimal design in random-effects regression models. *Biometrika*, 84(2):429–442, 1997.

[135] Noe B Mercado, Roland Zahn, Frank Wegmann, Carolin Loos, Abishek Chandrashekar, Jingyou Yu, Jinyan Liu, Lauren Peter, Katherine McMahan, Lisa H Tostanoski, et al. Single-shot ad26 vaccine protects against sars-cov-2 in rhesus macaques. *Nature*, 586(7830):583–588, 2020.

[136] Nicole Mideo, Samuel Alizon, and Troy Day. Linking within-and between-host dynamics in the evolutionary epidemiology of infectious diseases. *Trends in ecology & evolution*, 23(9):511–517, 2008.

[137] Risto Miikkulainen, Olivier Francon, Elliot Meyerson, Xin Qiu, Darren Sargent, Elisa Canzani, and Babak Hodjat. From prediction to prescription: evolutionary optimization of nonpharmaceutical interventions in the covid-19 pandemic. *IEEE Transactions on Evolutionary Computation*, 25(2):386–401, 2021.

[138] Philippe Moireau and Dominique Chapelle. Reduced-order unscented kalman filtering with application to parameter identification in large-dimensional systems. *ESAIM: Control, Optimisation and Calculus of Variations*, 17(2):380–405, 2011.

[139] Carl N Morris. Parametric empirical bayes inference: theory and applications. *Journal of the American statistical Association*, 78(381):47–55, 1983.

[140] César Muñoz-Fontela, William E Dowling, Simon GP Funnell, Pierre-S Gsell, A Ximena Riveros-Balta, Randy A Albrecht, Hanne Andersen, Ralph S Baric, Miles W Carroll, Marco Cavaleri, et al. Animal models for covid-19. *Nature*, 586(7830):509–515, 2020.

[141] Susan A Murphy, Kevin G Lynch, David Oslin, James R McKay, and Tom TenHave. Developing adaptive treatment strategies in substance abuse research. *Drug and alcohol dependence*, 88:S24–S30, 2007.

[142] Renato G Nascimento, Kajetan Fricke, and Felipe AC Viana. A tutorial on solving ordinary differential equations using python and hybrid physics-informed neural network. *Engineering Applications of Artificial Intelligence*, 96:103996, 2020.

[143] Nadège Néant, Guillaume Lingas, Quentin Le Hingrat, Jade Ghosn, Ilka Engelmann, Quentin Lepiller, Alexandre Gaymard, Virginie Ferré, Cédric Hartard, Jean-Christophe Plantier, et al. Modeling sars-cov-2 viral kinetics and association with mortality in hospitalized patients from the french covid cohort. *Proceedings of the National Academy of Sciences*, 118(8):e2017962118, 2021.

[144] THT Nguyen, M-S Mouksassi, Nicholas Holford, N Al-Huniti, I Freedman, Andrew C Hooker, J John, Mats O Karlsson, DR Mould, JJ Pérez Ruixo, et al. Model evaluation of continuous data pharmacometric models: metrics and graphics. *CPT: pharmacometrics & systems pharmacology*, 6(2):87–109, 2017.

[145] Martin A Nowak, Robert M May, and Roy M Anderson. The evolutionary dynamics of hiv-1 quasispecies and the development of immunodeficiency disease. *Aids*, 4(11):1095–1103, 1990.

[146] Joakim Nyberg, Caroline Bazzoli, Kay Ogungbenro, Alexander Aliev, Sergei Leonov, Stephen Duffull, Andrew C Hooker, and France Mentré. Methods and software tools for design evaluation in population pharmacokinetics–pharmacodynamics studies. *British journal of clinical pharmacology*, 79(1):6–17, 2015.

[147] Joseph O Ogutu and Hans-Peter Piepho. Regularized group regression methods for genomic prediction: Bridge, mcp, scad, group bridge, group lasso, sparse group lasso, group mcp and group scad. *BMC Proceedings*, 8(5):S7, 2014.

[148] Chloé Pasin, François Dufour, Laura Villain, Huilong Zhang, and Rodolphe Thiébaut. Controlling il-7 injections in hiv-infected patients. *Bulletin of mathematical biology*, 80:2349–2377, 2018.

[149] Alan S Perelson. Modelling viral and immune system dynamics. *Nature reviews immunology*, 2(1):28–36, 2002.

[150] Alan S Perelson and Ruian Ke. Mechanistic modeling of sars-cov-2 and other infectious diseases and the effects of therapeutics. *Clinical Pharmacology & Therapeutics*, 109(4):829–840, 2021.

[151] Alan S Perelson and Ruy M Ribeiro. Modeling the within-host dynamics of hiv infection. *BMC biology*, 11:1–10, 2013.

[152] Maya L Petersen, Steven G Deeks, Jeffrey N Martin, and Mark J van der Laan. History-adjusted marginal structural models for estimating time-varying effect modification. *American Journal of Epidemiology*, 166(9):985–993, 2007.

[153] José C Pinheiro and Douglas M Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4(1):12–35, 1995.

[154] Stanley A Plotkin. Updates on immunologic correlates of vaccine-induced protection. *Vaccine*, 38(9):2250–2257, 2020.

[155] Stanley A Plotkin and Peter B Gilbert. Nomenclature for immune correlates of protection after vaccination. *Clinical Infectious Diseases*, 54(11):1615–1617, 2012.

[156] AJ Pollard, O Launey, JD Lelievre, C Lacabaratz, M Gibani, and E Clutterbuck. Safety and immunogenicity of a two-dose heterologous ad26. zebov and mva-bn®-filo ebola vaccine regimen: a phase 2 randomised clinical study in europe (ebovac2). *Lancet Infectious Diseases*, 21(4), 2020.

[157] John S Preisser, Mary L Young, Daniel J Zaccaro, and Mark Wolfson. An integrated population-averaged approach to the design, analysis and sample size determination of cluster-unit trials. *Statistics in medicine*, 22(8):1235–1254, 2003.

[158] Ross L Prentice. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in medicine*, 8(4):431–440, 1989.

[159] Cécile Proust-Lima, Viviane Philipps, and Jean-François Dartigues. A joint model for multiple dynamic processes and clinical endpoints: Application to alzheimer's disease. *Statistics in Medicine*, 38(23):4702–4717, 2019.

[160] Hein Putter, SH Heisterkamp, JMA Lange, and F De Wolf. A bayesian approach to parameter estimation in hiv dynamical models. *Statistics in medicine*, 21(15):2199–2214, 2002.

[161] Stefan T Radev, Frederik Graw, Simiao Chen, Nico T Mutters, Vanessa M Eichel, Till Bärnighausen, and Ullrich Köthe. Outbreakflow: Model-based bayesian inference of disease outbreak dynamics with invertible neural networks and its application to the covid-19 pandemics in germany. *PLoS computational biology*, 17(10):e1009472, 2021.

[162] Adrian E Raftery and Le Bao. Estimating and projecting trends in hiv/aids generalized epidemics using incremental mixture importance sampling. *Biometrics*, 66(4):1162–1173, 2010.

[163] James O Ramsay and Xiaochun Li. Curve registration. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 60(2):351–363, 1998.

[164] Daniel B Reeves, Elizabeth R Duke, Sean M Hughes, Martin Prlic, Florian Hladik, and Joshua T Schiffer. Anti-proliferative therapy for hiv cure: a compound interest approach. *Scientific reports*, 7(1):4011, 2017.

[165] Jason A Regules, John H Beigel, Kristopher M Paolino, Jocelyn Voell, Amy R Castellano, Zonghui Hu, Paula Muñoz, James E Moon, Richard C Ruck, Jason W Bennett, et al. A recombinant vesicular stomatitis virus ebola vaccine. *New England Journal of Medicine*, 376(4):330–341, 2017.

[166] Bastien Reyné, Mircea T Sofonea, and Samuel Alizon. Modelling long-term covid-19 hospital admission dynamics using immune protection waning data. *medRxiv*, pages 2022–05, 2022.

[167] John Ritz and Donna Spiegelman. Equivalence of conditional and marginal regression models for clustered and longitudinal data. *Statistical Methods in Medical Research*, 13(4):309–323, 2004.

[168] James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, pages 550–560, 2000.

[169] Libin Rong and Alan S Perelson. Modeling hiv persistence, the latent reservoir, and viral blips. *Journal of theoretical biology*, 260(2):308–331, 2009.

[170] IM Rouzine, RA Sergeev, and AI Glushtsov. Two types of cytotoxic lymphocyte regulation explain kinetics of immune response to human immunodeficiency virus. *Proceedings of the National Academy of Sciences*, 103(3):666–671, 2006.

[171] Olli Saarela, Elja Arjas, David A Stephens, and Erica EM Moodie. Predictive bayesian inference and dynamic treatment regimes. *Biometrical Journal*, 57(6):941–958, 2015.

[172] Adeline Samson, Marc Lavielle, and France Mentré. Extension of the saem algorithm to left-censored data in nonlinear mixed-effects model: Application to hiv dynamics model. *Computational Statistics & Data Analysis*, 51(3):1562–1574, 2006.

[173] Joseph L Schafer. *Analysis of incomplete multivariate data*. CRC press, 1997.

[174] Joshua T Schiffer. Correlates of protection via modeling. *Nature Computational Science*, 2(3):140–141, 2022.

[175] Joshua T Schiffer, David A Swan, Amalia Magaret, Lawrence Corey, Anna Wald, Joachim Ossig, Helga Ruebsamen-Schaeff, Susanne Stoelben, Burkhard Timmler, Holger Zimmermann, et al. Mathematical modeling of herpes simplex virus-2 suppression with pritelivir predicts trial outcomes. *Science translational medicine*, 8(324):324ra15–324ra15, 2016.

[176] Christian Selinger and Samuel Alizon. Reconstructing contact network structure and cross-immunity patterns from multiple infection histories. *PLoS Computational Biology*, 17(9):e1009375, 2021.

[177] Christian Selinger, Marc Choisy, and Samuel Alizon. Predicting covid-19 incidence in french hospitals using human contact network analytics. *International Journal of Infectious Diseases*, 111:100–107, 2021.

[178] Gilles-Eric Séralini, Dominique Cellier, and Joël Spiroux de Vendomois. New analysis of a rat feeding study with a genetically modified maize reveals signs of hepatorenal toxicity. *Archives of environmental contamination and toxicology*, 52:596–602, 2007.

[179] Lewis B Sheiner, Barr Rosenberg, and Vinay V Marathe. Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *Journal of pharmacokinetics and biopharmaceutics*, 5:445–479, 1977.

[180] Lewis B Sheiner, Barr Rosenberg, and Kenneth L Melmon. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Computers and Biomedical Research*, 5(5):441–459, 1972.

[181] Georgi Shukarev, Benoit Callendret, Kerstin Luhn, Macaya Douoguih, and EBOVAC1 consortium. A two-dose heterologous prime-boost vaccine regimen eliciting sustained immune responses to ebola zaire could support a preventive strategy for future outbreaks. *Human vaccines & immunotherapeutics*, 13(2):266–270, 2017.

[182] Chris J Skinner. Inverse probability weighting for clustered nonresponse. *Biometrika*, 98(4):953–966, 2011.

[183] Mircea T Sofonea, Samuel Alizon, and Yannis Michalakis. From within-host interactions to epidemiological competition: a general model for multiple infections. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1675):20140303, 2015.

[184] Mircea T Sofonea, Bénédicte Roquebert, Vincent Foulongne, David Morquin, Laura Verdurme, Sabine Trombert-Paolantoni, Mathilde Roussel, Jean-Christophe Bonetti, Judith Zerah, Stéphanie Haim-Boukobza, et al. Analyzing and modeling the spread of sars-cov-2 omicron lineages ba. 1 and ba. 2, france, september 2021–february 2022. *Emerging Infectious Diseases*, 28(7):1355, 2022.

[185] Robert F Stengel and Raffaele Ghigliazza. Stochastic optimal therapy for enhanced immune response. *Mathematical biosciences*, 191(2):123–142, 2004.

[186] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[187] David A Swan, Chloe Bracis, Holly Janes, Mia Moore, Laura Matrajt, Daniel B Reeves, Eileen Burns, Deborah Donnell, Myron S Cohen, Joshua T Schiffer, et al. Covid-19 vaccines that reduce symptoms but do not block infection need higher coverage and faster rollout to achieve population impact. *Scientific reports*, 11(1):15531, 2021.

[188] Bachirou O Taddé, Hélène Jacqmin-Gadda, Jean-François Dartigues, Daniel Commenges, and Cécile Proust-Lima. Dynamic modeling of multivariate dimensions and their temporal relationships using latent processes: Application to alzheimer's disease. *Biometrics*, 76(3):886–899, 2020.

[189] Luís Tarrataca, Claudia Mazza Dias, Diego Barreto Haddad, and Edilson Fernandes De Arruda. Flattening the curves: on-off lock-down strategies for covid-19 with an application to brazil. *Journal of mathematics in industry*, 11:1–18, 2021.

[190] Yuichi Tatsukawa, Md Rajib Arefin, Kazuki Kuga, and Jun Tanimoto. An agent-based nested model integrating within-host and between-host mechanisms to predict an epidemic. *PloS one*, 18(12):e0295954, 2023.

[191] Alexandra Tauzin, Manon Nayrac, Mehdi Benlarbi, Shang Yu Gong, Romain Gasser, Guillaume Beaudoin-Bussières, Nathalie Brassard, Annemarie Laumaea, Dani Vézina, Jérémie Prévost, et al. A single dose of the sars-cov-2 vaccine bnt162b2 elicits fc-mediated antibody effector functions and t cell responses. *Cell host & microbe*, 29(7):1137–1150, 2021.

[192] PREVAC Study Team. Randomized trial of vaccines for zaire ebola virus disease. *New England Journal of Medicine*, 387(26):2411–2424, 2022.

[193] Rodolphe Thiébaut, Philippe Morlat, Hélène Jacqmin-Gadda, Didier Neau, Patrick Mercié, François Dabis, Geneviève Chêne, et al. Clinical progression of hiv-1 infection according to the viral response during the first year of antiretroviral treatment. *Aids*, 14(8):971–978, 2000.

[194] Robin N Thompson, Jake E Stockwin, Rolina D van Gaalen, Jonny A Polonsky, Zhian N Kamvar, P Alex Demarsh, Elisabeth Dahlqwist, Siyang Li, Eve Miguel, Thibaut Jombart, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29:100356, 2019.

[195] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

[196] Christoffer W Tornøe, Rune V Overgaard, Henrik Agersø, Henrik A Nielsen, Henrik Madsen, and E Niclas Jonsson. Stochastic differential equations in nonmem®: implementation, application, and comparison with ordinary differential equations. *Pharmaceutical research*, 22:1247–1258, 2005.

[197] Adam Trickey, Caroline A Sabin, Greer Burkholder, Heidi Crane, Antonella d'Arminio Monforte, Matthias Egger, M John Gill, Sophie Grabar, Jodie L Guest, Inma Jarrin, et al. Life expectancy after 2015 of adults with hiv on long-term antiretroviral therapy in europe and north america: a collaborative analysis of cohort studies. *The Lancet HIV*, 10(5):e295–e307, 2023.

[198] Nathan Trouvain, Luca Pedrelli, Thanh Trung Dinh, and Xavier Hinaut. ReservoirPy: An efficient and user-friendly library to design echo state networks. In *Artificial Neural Networks and Machine Learning – ICANN 2020*, pages 494–505. Springer International Publishing, 2020.

[199] Anastasios A Tsiatis. *Semiparametric theory and missing data*. Springer, 2006.

[200] Mark J Van der Laan and Maya L Petersen. Causal effect models for realistic individualized treatment and intention to treat rules. *The international journal of biostatistics*, 3(1), 2007.

[201] Tyler J VanderWeele. Surrogate measures and consistent surrogates. *Biometrics*, 69(3):561–565, 2013.

[202] Sarah C Vick, Marie Frutoso, Florian Mair, Andrew J Konecny, Evan Greene, Caitlin R Wolf, Jennifer K Logue, Nicholas M Franko, Jim Boonyaratanakornkit, Raphael Gottardo, et al. A regulatory t cell signature distinguishes the immune landscape of covid-19 patients from those with other respiratory infections. *Science advances*, 7(46):eabj0274, 2021.

[203] Alexander M Walker. Confounding by indication. *Epidemiology*, 7(4):335–336, 1996.

[204] L Wang, Jiguo Cao, James O Ramsay, DM Burger, CJL Laporte, and Jürgen K Rockstroh. Estimating mixed-effects differential equation models. *Statistics and Computing*, 24:111–121, 2014.

[205] Xiping Wei, Sajal K Ghosh, Maria E Taylor, Victoria A Johnson, Emilio A Emini, Paul Deutsch, Jeffrey D Lifson, Sebastian Bonhoeffer, Martin A Nowak, Beatrice H Hahn, et al. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*, 373(6510):117–122, 1995.

[206] Sue J Welham, Brian R Cullis, Michael G Kenward, and Robin Thompson. A comparison of mixed model splines for curve fitting. *Australian & New Zealand Journal of Statistics*, 49(1):1–23, 2007.

[207] GB Wetherill. Bayesian sequential analysis. *Biometrika*, 48(3/4):281–292, 1961.

[208] Michael T White, Jamie T Griffin, Onome Akpogheneta, David J Conway, Kwadwo A Koram, Eleanor M Riley, and Azra C Ghani. Dynamics of the antibody response to plasmodium falciparum infection in african children. *The Journal of infectious diseases*, 210(7):1115–1122, 2014.

[209] Hulin Wu, Tao Lu, Hongqi Xue, and Hua Liang. Sparse additive ordinary differential equations for dynamic gene regulatory network modeling. *Journal of the American Statistical Association*, 109(506):700–716, 2014.

[210] Hulin Wu and Jin-Ting Zhang. *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*. John Wiley & Sons, 2006.

[211] Yanxun Xu, Peter Müller, Abdus S Wahed, and Peter F Thall. Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *Journal of the American Statistical Association*, 111(515):921–950, 2016.

[212] Reza Yaesoubi, Joshua Havumaki, Melanie H Chitwood, Nicolas A Menzies, Gregg Gonsalves, Joshua A Salomon, A David Paltiel, and Ted Cohen. Adaptive policies to balance health benefits and economic costs of physical distancing interventions during the covid-19 pandemic. *Medical Decision Making*, 41(4):386–392, 2021.

[213] Safoora Yousefi, Fatemeh Amrollahi, Mohamed Amgad, Chengliang Dong, Joshua E Lewis, Congzheng Song, David A Gutman, Sameer H Halani, Jose Enrique Velazquez Vega, Daniel J Brat, et al. Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. *Scientific reports*, 7(1):11707, 2017.

[214] Jingyou Yu, Lisa H Tostanoski, Lauren Peter, Noe B Mercado, Katherine McMahan, Shant H Mahrokhian, Joseph P Nkolola, Jinyan Liu, Zhenfeng Li, Abishek Chandrashekar, et al. Dna vaccine protection against sars-cov-2 in rhesus macaques. *Science*, 369(6505):806–811, 2020.

[215] Zhuo Yu and Mark J van der Laan. Construction of counterfactuals and the g-computation formula. *bepress*, 2002.

[216] Yao Zhang and Mihaela van der Schaar. Gradient regularized v-learning for dynamic treatment regimes. *Advances in Neural Information Processing Systems*, 33:2245–2256, 2020.

[217] Zizhen Zhang, Zhiyuan Wu, Hang Zhang, and Jiahai Wang. Meta-learning-based deep reinforcement learning for multiobjective optimization problems. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

Habilitation à Diriger des Recherches de l'Université de Bordeaux

Résumé

Ce manuscrit offre un résumé des recherches que j'ai menées ou encadrées, entre 2010 et 2024, dans le champ de la biostatistique appliquée à la modélisation mathématique des maladies infectieuses. Mon exposé se structure en trois parties : le développement méthodologique pour l'utilisation de modèles mécanistes, leur application à la modélisation intra-hôte de la dynamique virale et du système immunitaire, et à l'évaluation en population des interventions de contrôle des épidémies. Les applications principales concernent le VIH, Ebola, le virus Nipah et le Sars-CoV-2. Je conclue en présentant mes perspectives de recherche futures et ma vision des progrès significatifs attendus de ces modèles.

Mots-clefs Biostatistique; Causalité; Ebola; HIV; Maladies infectieuses; Inférence; Modèles mécanistes; Modèles à effets mixtes non linéaires; SARS-CoV-2.

Abstract

This manuscript provides a summary of the research I have conducted or supervised between 2010 and 2024, in the field of biostatistics applied to the mathematical modeling of infectious diseases. My presentation is structured into three parts: the methodological development for the use of mechanistic models, their application to the intra-host modeling of viral dynamics and the immune system, and to the population-level evaluation of epidemic control interventions. The primary applications include HIV, Ebola, the Nipah virus, and Sars-CoV-2. I conclude by presenting my future research directions and my vision of the significant advances expected from these models.

Keywords Biostatistics; Causality; Ebola; HIV; Infectious diseases; Inference; Mechanistic models; Non-linear Mixed effects models; SARS-CoV-2.

Laboratory: Inria Bordeaux Sud-Ouest; Inserm Bordeaux Population Health U1219; 146 Rue Leo Saignat 33000 Bordeaux, France

université de BORDEAUX