



**HAL**  
open science

## Analyses quantitatives par temps de crise sanitaire

Mircea T Sofonea

► **To cite this version:**

Mircea T Sofonea. Analyses quantitatives par temps de crise sanitaire. Santé publique et épidémiologie. Sorbonne Université - Faculté de Médecine, 2024. tel-04752423

**HAL Id: tel-04752423**

**<https://hal.science/tel-04752423v1>**

Submitted on 24 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# Analyses quantitatives par temps de crise sanitaire

modèles pour l'étude et le contrôle  
du SARS-CoV-2 en France

Mircea T. SOFONEA



## *Habilitation à diriger des recherches*

mémoire déposé le 1<sup>er</sup> septembre 2023 et soutenu publiquement le 11 mars 2024  
à Paris (XIII<sup>ème</sup>) devant le jury composé de

Vincent CALVEZ,	DR CNRS,	rapporteur
Vittoria COLIZZA,	DR Inserm & Georgetown Univ. (États-Unis),	examinatrice
Dominique COSTAGLIOLA,	DR Inserm,	présidente du jury
Antoine FLAHAULT,	PU Univ. de Genève (Suisse),	rapporteur
Solen KERNÉIS,	PU-PH Univ. Paris Cité & AP-HP,	examinatrice
Mahmoud ZUREIK,	PU-PH Univ. Versailles - St-Quentin-en-Yvelines & ANSM,	rapporteur



*À ma mère, à mon père et à mon fils.*



*Vreme trece, vreme vine,  
Toate-s vechi și nouă toate ;  
Ce e rău și ce e bine  
Tu te-ntreabă și socoate ;  
Nu spera și nu ai teamă,  
Ce e val ca valul trece ;  
De te-ndeamnă, de te cheamă,  
Tu rămâi la toate rece. [...]*

Le temps passe, le temps vient,  
Tout est ancien, tout est nouveau ;  
Ce qui est mal, ce qui est bien  
Pèse et médite à tout propos ;  
N'espère pas et n'aie pas peur,  
Ce qui est flot en flot s'en va,  
Si on te mande ou on te leurre.  
À toute chose reste froid. [...]

*Glossă, Mihai EMINESCU, 1883.*



Rue Jean-de-Beauvais,  
Paris 5<sup>e</sup>, oct. 2022

# Table des matières

<b>1</b>	<b>Curriculum vitae &amp; travaux</b>	<b>2</b>
1.1	Parcours & activités	4
1.2	Travaux	14
<b>2</b>	<b>Encadrement &amp; activités connexes</b>	<b>26</b>
2.1	Stages de master	28
2.2	Thèses d'exercice	28
2.3	Thèses de doctorat	29
2.4	Contrats d'ingénieurs & post-doctorants	33
<b>3</b>	<b>Mémoire : synthèse &amp; perspectives de recherche</b>	<b>36</b>
3.1	Synthèse des travaux réalisés	38
3.2	Perspectives à court terme	60
3.3	Perspectives à long terme	66
<b>4</b>	<b>Annexes</b>	<b>80</b>
4.1	Publications sélectionnées	82
4.2	Diplôme	127
	<i>Bibliographie</i>	130
4.3	Remerciements	142

À l'attention du lecteur du document numérique : les éléments colorés (cliquables) en **rouge fuchsia** renvoient vers une autre page, une figure ou une note de bas de page de ce même document, tandis que ceux colorés en **bleu glauque** renvoient vers une page web. Les citations (lesquelles renvoient vers la bibliographie), ainsi que les marges de tableau sont **grisées** afin d'alléger la lecture.

3

**Mémoire : synthèse &  
perspectives de  
recherche**

## 3.1 Synthèse des travaux réalisés

Par souci d'homogénéité et de concision, seuls mes travaux portant sur la pandémie de COVID-19 sont présentés dans cette section. Depuis ma thèse – et y compris au cours de la pandémie – ma recherche a aussi concerné la modélisation des infections multiples (Gallet et al., 2018, Sofonea, 2019, Zwart et al., 2021), l'antibiorésistance (Djidjou-Demasse et al., 2021, Djidjou-Demasse et al., 2023), les papillomavirus humains (Alizon et al., 2019, Beneteau et al., 2021b), et l'activité pré-pandémique des soins critiques en France (Boulet et al., 2023).

### 3.1.1 De la modélisation et de sa place

Fin décembre 2019, un cluster de cas de pneumonies atypiques d'étiologie inconnue est détecté à Wuhan, en Chine centrale (Tan et al., 2020; Zhu et al., 2020). En l'espace d'un mois, le coronavirus SARS-CoV-2 sera séquencé (Wu et al., 2020; Gorbalenya et al., 2020) et les cinétiques de l'histoire naturelle de la maladie respiratoire qu'il cause, le COVID-19, quantifiées (Li et al., 2020). Au même moment, des approches statistiques essayant d'inférer la taille de l'épidémie à partir du nombre limité de cas détectés en dehors de Chine sont mises en ligne (Imai et al., 2020). Deux mois plus tard, l'Organisation Mondiale de la Santé (OMS) déclare le COVID-19 comme pandémie (11/03/2020), et des simulations de collègues de l'*Imperial College London* portant sur les leviers non pharmaceutiques de prévention explorent des scénarios de conséquences sanitaires au Royaume-Uni avec un horizon de près de deux ans (Ferguson et al. 2020, 16/03/2020). Dépassant le cadre de la science et de la médecine, l'impact de la modélisation en santé publique sur les plans politique, socio-économique et médiatique est alors sans précédent (Adam, 2020).

Tout au long de la pandémie, la modélisation a été utilisée pour caractériser les risques associés au COVID-19, prévoir l'impact sur le système de santé (Paireau et al., 2022), aider à la décision (Currie et al., 2020), guider le déploiement optimal des vaccins (Tran Kiem et al., 2021), et de nombreux aspects de notre vie avec le virus (Haug et al., 2020). Pourtant, compte tenu de l'urgence et des enjeux de ces trois années écoulées, l'attention s'est quasi-exclusivement portée sur les résultats projectifs (plus exactement la fraction la plus anxiogène d'entre eux), laissant relativement peu de place pour expliciter l'approche elle-même et discuter de son rôle pléiotrope dans la crise sanitaire – en distinguant le souhaitable du possible et de l'effectif. La présente section constitue une opportunité pour ce faire, en reprenant largement les publications (Djidjou-Demasse et al. 2020b, Sofonea et al. 2022a) qui s'inscrivaient dans cette démarche.

En santé humaine comme dans d'autres sujets d'étude scientifique (Cohen, 2004; Servedio et al., 2014), les modèles visent à fournir une représentation simplifiée du phénomène étudié, en se concentrant sur le sous-ensemble de propriétés et de processus considérés comme des moteurs

essentiels. La structure de ces modèles résulte donc de choix motivés par des connaissances, des données et des observations antérieures et d'un ensemble d'hypothèses fondées sur notre compréhension du phénomène (Goldstein, 2018). Équipés d'un formalisme, d'hypothèses de travail et désormais d'un corpus théorique important, les modèles (qu'ils soient mathématiques (Keeling and Rohani, 2008; Diekmann et al., 2013), statistiques (Rothman and Greenland, 2005; Bjornstad, 2022) ou computationnels (Arifin et al., 2016; Gaudou et al., 2020)) fournissent un cadre adapté à l'étude des maladies infectieuses, leur surveillance et leur contrôle. Il est important de noter que même les analyses quantitatives les plus communes (telle une régression linéaire) relèvent de la modélisation, bien que leurs hypothèses ne soient pas toujours explicites – ce qui au passage expose au risque de ne pas les vérifier – (p.ex. la linéarité, l'absence d'auto-corrélation, l'homoscédasticité, la faible multicolinéarité, la normalité).

Un modèle épistémologiquement pertinent est manipulable, génératif (de preuves de concepts, de quantifications ou de conjectures sur le passé, les contrefactuels, ou le futur, les projections) et réfutable, et relève généralement d'un des quatre grands objectifs suivants – parmi ceux listés par la fig. 3.1 – et ci-après illustrés par quatre des cinq travaux sélectionnés en annexe de ce document :

- la **description**, *e.g.* celle de la cinétique effective de remplacement des variants Alpha, Beta et Gamma du SARS-CoV-2 par le variant Delta (juin 2021) à partir de RT-qPCR de criblages (Alizon et al., 2021), *cf.* p. 92,
- l'**évaluation**, *e.g.* celle de l'efficacité du premier confinement national en termes de réduction du nombre de reproduction  $\mathcal{R}_t$  (Sofonea et al., 2021c), *cf.* p.84,
- la **compréhension**, *e.g.* en mettant en évidence que la propagation du variant Omicron BA.2 ne peut expliquer à elle seule le rebond épidémique à l'origine de la sixième vague (Sofonea et al., 2022b), *cf.* p. 116,
- l'**anticipation**, *e.g.* celle des besoins en soins critiques par projection des lits occupés à partir de l'épidémiologie en population générale (Sofonea and Alizon, 2021), *cf.* p. 98.

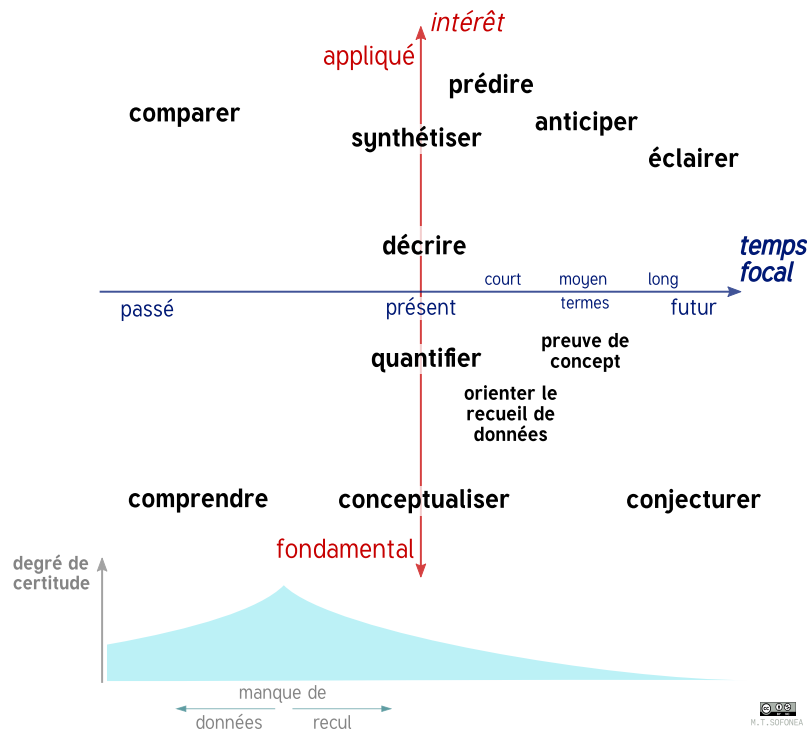


FIGURE 3.1 – **Diversité des questions abordées par les modèles épidémiologiques.**

Les objectifs de la modélisation peuvent être décrits selon deux axes. Le premier porte sur la temporalité du phénomène par rapport à l'étude : immédiate (*nowcasting*), rétrospective ou prospective. Le second s'identifie à l'intérêt, fondamental ou appliqué, des résultats apportés par le modèle. La disposition absolue des objectifs dans le repère ainsi que la distribution du degré de certitude est arbitraire et n'a qu'une valeur indicative. En première approximation, la modélisation peut être utilisée aussi bien pour comprendre rétrospectivement la dynamique d'une épidémie, quantifier une propagation en cours ou anticiper la trajectoire à venir.

Au-delà des divergences d'objectifs, les modèles se déclinent aussi selon la discipline focale (microbiologie, génomique/biologie évolutive, infectiologie/immunologie, santé publique/épidémiologie), l'échelle étudiée (intra-hôte, sphère d'interaction, populations) et la méthodologie (phénoménologique – statistique, apprentissage machine –, ou mécanistique – compartimental ou individu-centré).

Il existe plusieurs raisons de recourir à la modélisation lorsque l'on étudie les maladies infectieuses, pour leur compréhension, leur surveillance ou leur contrôle. Tout d'abord, l'intrication des différents processus (physiopathologie, transmission, politiques sanitaires) est un obstacle à l'intuition et à l'argumentation qualitative. Deuxièmement, il est rarement possible de mesurer l'effet des interventions par observation directe. Contrairement aux prises en charge médicales qui peuvent être évaluées au niveau du patient et comparées, de façon éthique, entre des groupes de patients, les interventions relatives à l'épidémie elle-même ne peuvent généralement être évaluées qu'à l'échelle populationnelle, en vie réelle. Dans ce contexte, la modélisation est le seul moyen de comparer les scénarios d'évolution possibles. Illustration plus proximale, la valeur prédictive positive (VPP) d'un test de dépistage nous rappelle que la confiance accordée à un résultat individuel ne dépend pas seulement des caractéristiques du test mais aussi de la prévalence de l'infection dans l'ensemble de la population. Une troisième



difficulté réside dans la nature non linéaire de la dynamique des maladies transmissibles, laquelle peut, d'une part, être bruitée par la stochasticité (Lloyd-Smith et al. 2005) et, d'autre part, engendrer des auto-corrélations spatiales et temporelles qu'il convient de prendre en compte dans les analyses inférentielles (Gaudart et al., 2021). Un quatrième défi est représenté par la combinaison de biais omniprésents dans ces circonstances. Il s'agit aussi bien de biais de sélection induits par la censure temporelle à droite (Britton and Scalia Tomba, 2019; Russell et al., 2020) (pouvant aller jusqu'à inverser la chronologie de la chaîne de transmission (Geismar et al., 2023)), par la variabilité des symptômes (Lipsitch et al., 2015) ou de la susceptibilité (Monge et al., 2023), de confusion (von Kügelgen et al., 2021), que de biais de classement (Surkova et al., 2020). Ces obstacles ne peuvent être contournés qu'avec une méthodologie rigoureuse et transparente, parfois même encore discutée, sous peine de résultats erronés, de crise de reproductibilité, de coût sanitaire et de perte de confiance en la recherche biomédicale (Supplisson and Sofonea, 2023). En particulier, la propagation de l'incertitude, l'analyse de sensibilité, le calcul d'intervalles statistiques permettent de fournir une mesure de la confiance/crédibilité sans laquelle l'estimation ponctuelle serait insuffisamment informative (Zelner et al., 2021).

La modélisation en épidémiologie des maladies infectieuses repose aujourd'hui sur une variété de méthodes basées sur des décennies de développements à l'interface de la démographie, de la médecine et de la biologie (Foppa, 2016). La grande majorité des modèles adopte l'historique approche compartimentale, laquelle partitionne la population en groupes considérés comme homogènes d'un point de vue clinico-épidémiologique. L'exemple paradigmatique est le modèle *SIR* qui décrit les individus comme étant d'abord sensibles à l'infection (*S*), puis infectés par la maladie et contagieux (*I*), et enfin retirés de la transmission par la guérison ou le décès (*R*) (Kermack and McKendrick, 1927). La simplicité d'implémentation, de simulation et la grande flexibilité de cette approche expliquent l'existence de modèles compartimentaux pour de nombreuses maladies infectieuses et situations de contrôle (Keeling and Rohani, 2008). Ils permettent en effet de rendre compte de façon parcimonieuse (i.e. *a priori* avec peu de degrés de liberté) de la dynamique non linéaire tout en identifiant (au moins partiellement) les paramètres à des quantités mesurables en vie réelle et même de produire d'élégants résultats analytiques (Diekmann et al., 2013), dont l'incontournable l'expression du nombre de reproduction de base en fonction des paramètres clés. Malgré la simplicité structurelle de la modélisation compartimentale, la littérature de méta-modélisation a confirmé la robustesse et la pertinence sanitaire des résultats produits par cette méthodologie, au moins au démarrage d'une épidémie (Trapman et al., 2016). Déterministes la plupart du temps (la loi des grands nombres aidant), ces modèles peuvent aisément se décliner en version stochastique lorsque la variabilité inter-individuelle est importante (Elie et al., 2022b) et que l'on souhaite quantifier l'impact de cette hétérogénéité sur la dynamique épidémiologique, notamment dans des contextes où le nombre de cas est faible (p.ex. initiation ou extinction d'une épidémie (Beneteau et al., 2021a), y compris localement). Lorsque les sources d'hétérogénéités (compartimentales, spatiales, temporelles) et les problèmes d'identifiabilité se font trop nombreux, deux méthodologies épistémologiquement opposées prennent souvent le relais : d'un côté la réduction par approche essentiellement phé-

noménologique reposant sur des techniques statistiques et d'apprentissage machine, parfois combinées linéairement (Paireau et al., 2022), et de l'autre l'individualisation par les simulations orientées agents, où chaque individu interagit explicitement avec les autres (Thomine et al., 2021).

Les trois approches majeures présentent des avantages et des inconvénients distincts et aucune ne permet de maximiser simultanément le réalisme, l'intelligibilité et la généralité du modèle dont la construction résulte donc d'un compromis arbitraire. Tous les aspects d'un phénomène aussi riche structurellement et dynamiquement qu'une pandémie ne peuvent donc être abordés par un unique modèle. Bien qu'un cadre de travail puisse éventuellement répondre à plusieurs questions connexes, sa transposition à une autre temporalité, une autre échelle spatiale voire à une autre maladie infectieuse (p.ex. la 'triple épidémie' de viroses respiratoires, cf. p. 60), ne peut se faire qu'au moyen d'ajustements importants qui sont au cœur de l'analyse critique de cette littérature située aux interfaces. Dès lors, une décision relative à la santé publique ne saurait être prise sur la base d'une simulation isolée, mais devrait reposer sur un ensemble de résultats congruents issus de modèles indépendants et méthodologiquement complémentaires, pondérés par leur adéquation aux données disponibles – voir la fig. 3.2, qui rappelle que les modèles doivent dès leur conception être vus comme des moyens et non comme des fins.

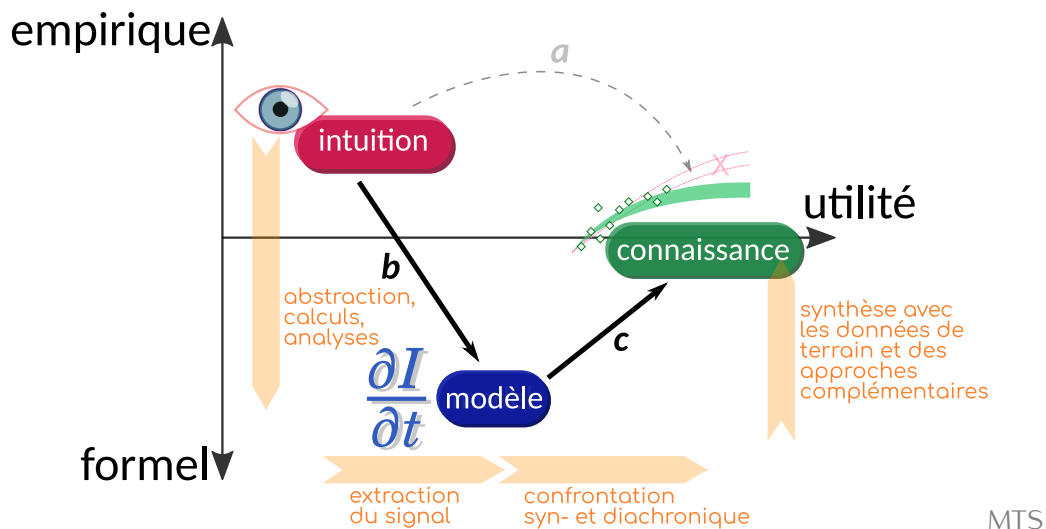


FIGURE 3.2 – **Les modèles sont des étapes intermédiaires souvent nécessaires entre l'observation et la connaissance.**

Idéalement l'étude d'un phénomène devrait transformer directement (ou tout au plus via l'intuition) les observations empiriques en connaissances établies. Dans le cas d'une épidémie, cette voie (a) est difficile à suivre en raison de nombreuses limitations éthiques, spatio-temporelles, logistiques ou de processus simultanés. La modélisation constitue un pas de côté pour extraire du signal à partir de données d'observation (modulo certaines hypothèses) (étape b), permettant ainsi une manipulation quantitative dans un espace formalisé. Les comparaisons longitudinales (performance du modèle dans le passé) et transversales (situations ici/ailleurs) avec les données de terrain ainsi que les explorations systématiques du modèle (analyses de sensibilité) permettent dans un second temps d'établir les connaissances recherchées (étape c).

## 3.1.2 Un modèle central

### 3.1.2.1 Motivations et spécificités

Fin mars 2020, une des priorités nationales était l'évitement, ou à défaut l'anticipation, de la saturation des services hospitaliers de soins critiques\*. La modélisation apparaissait être un outil de choix pour contribuer à l'effort collectif. Il fallait tout d'abord identifier les problèmes inhérents à la flambée d'une maladie infectieuse inconnue – le manque de recul, le manque de données (en particulier nationales), les biais (notamment de sélection), et la non stationnarité spatio-temporelle de plusieurs processus à l'œuvre dans l'épidémiologie elle-même et son suivi. Ce constat a motivé le choix d'une approche parcimonieuse telle qu'imaginée par la fig. 3.3, c'est-à-dire privilégier un faible nombre de degrés de liberté† pour viser une anticipation robuste du temps long‡, plutôt qu'un ajustement immédiat mais fragile à l'arrivée de nouvelles données.

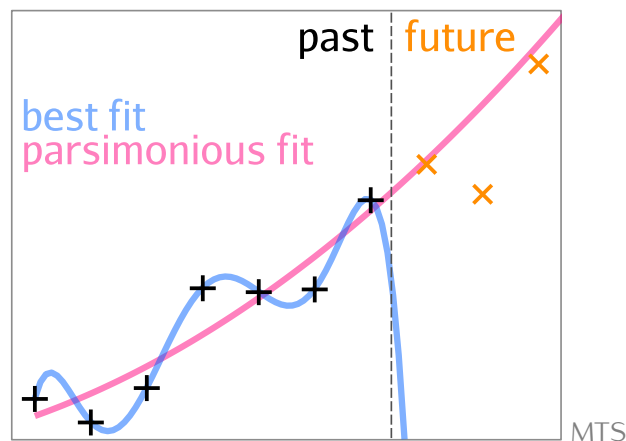


FIGURE 3.3 – **Illustration fictive de la motivation d'une modélisation parcimonieuse.** Une robustesse future est préférable à un sur-ajustement présent : ici, un polynôme de degré 6 peut passer exactement par 7 points, mais échoue à capturer la tendance à venir, contrairement à une simple parabole qui ne fait qu'approcher les données connues.

Au-delà de l'anticipation de la date et de la hauteur du pic hospitalier, il était souhaitable que le modèle puisse aussi apporter des éléments de réponse à d'autres questions soulevées à l'époque : l'immunité collective allait-elle bientôt être atteinte ? le confinement était-il efficace ? aurait-on pu réagir autrement ? Il était donc essentiel de pouvoir explorer numériquement des scénarios alternatifs (contrefactuels) et donc de disposer de paramètres libres du modèles identifiables aux mesures sanitaires (en intensité et en timing), une capacité offerte par la modélisation mécanistique§. Pour compenser le sous-ajustement de temps court, il convenait

\*. Dont la capacité nationale était estimée, fin 2019, à 5 080 lits (réanimation, soins intensifs et surveillance continue) (Lefrant et al., 2021).

†. En pratique il s'agira de moins de 10 paramètres libres, par fenêtre temporelle.

‡. En prévision des futures vagues qui faisaient [peu de doute pour nous](#).

§. C'est-à-dire qui reproduit le phénomène focal – ici l'occupation des lits de soins critiques – par l'interaction entre entités structurellement emboîtées – la transmission communautaire. Par opposition aux modèles phénoménologiques (statistiques, apprentissage machine) qui reproduisent les patrons eux-mêmes sans identification de processus.

de capturer précisément la cinétique de l'histoire naturelle de la maladie, dont les premiers éléments étaient déjà connus en février 2020 (Bouadma et al., 2020; Linton et al., 2020; Nishiura et al., 2020) mais qui ne sera précisément résolue que plusieurs mois plus tard<sup>\*</sup>, voir fig. 3.4.

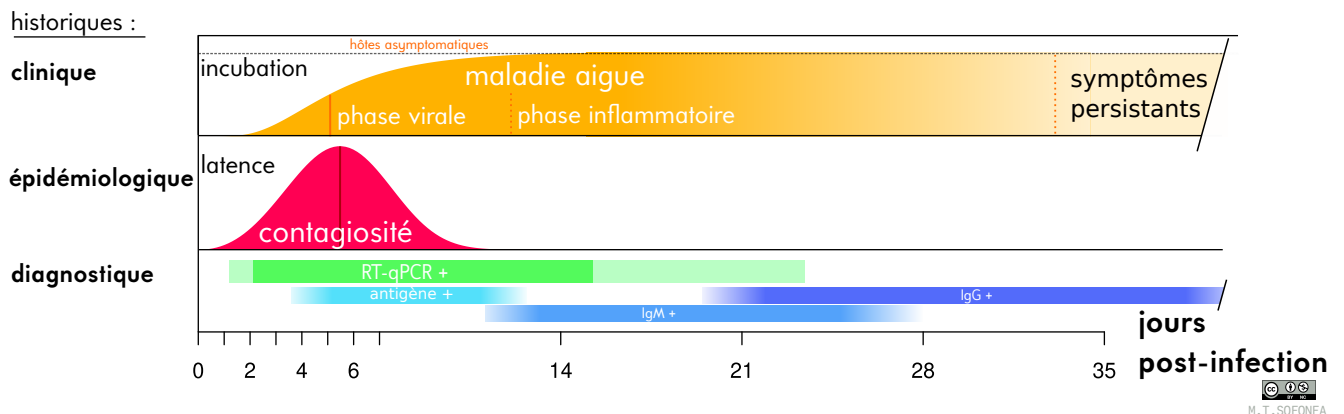


FIGURE 3.4 – **Histoire semi-quantitative clinique, épidémiologique et diagnostique du COVID-19.**

1) Clinique : la fonction de répartition de la période d'incubation (temps écoulé entre l'infection et l'apparition des premiers symptômes) est issue de (McAloon et al., 2020) (la barre verticale représente la médiane), le découpage des phases cliniques est abordé dans (Bouadma et al., 2020; Polak et al., 2020; Nalbandian et al., 2021) et positionné relativement au temps d'incubation médian. La fraction asymptomatique est celle estimée par (Byambasuren et al., 2020).

2) La contagiosité est donnée par la densité de probabilité du temps de génération, dont les paramètres ont été estimés par (Ferretti et al., 2020) (la barre verticale représente la médiane). Le délai séparant l'infection du début de la contagiosité est appelé latence.

3) La cinétique de positivité des tests nasopharyngés par RT-qPCR est inférée par (Hellewell et al., 2021) (plus de 50 % des cas sont positifs dans la bande verte centrale, plus de 5 % dans les bandes vert clair périphériques). Les positivités des tests antigéniques et sérologiques (immunoglobulines (Ig) M et G) sont uniquement qualitatives, d'après (Mercer and Salit, 2021).

Qu'il s'agisse de la contagiosité (caractérisée par l'intervalle sériel), ou du risque d'hospitalisation, leurs cinétiques apparaissent dépendantes du temps écoulé depuis le début de l'infection. Or les modèles compartimentaux les plus simples (des systèmes de quelques équations différentielles ordinaires) ne peuvent structurellement reproduire que des cinétiques sans mémoire (i.e. la variation instantanée ne dépend que du présent, propriété dite de Markov). Les temps d'attente des transitions inter-compartiments sont alors distribués exponentiellement, engendrant des flux d'admissions plus étalés dans le temps, empêchant de reproduire le pic en ligne de mire (celui constitué par la cohorte d'individus infectés les quelques jours qui ont précédé le confinement), comme l'illustre la fig. 3.5.

\*. Sans compter la quantification fine et la mise à jour variant après variant, voir notamment (Geismar et al., 2023).

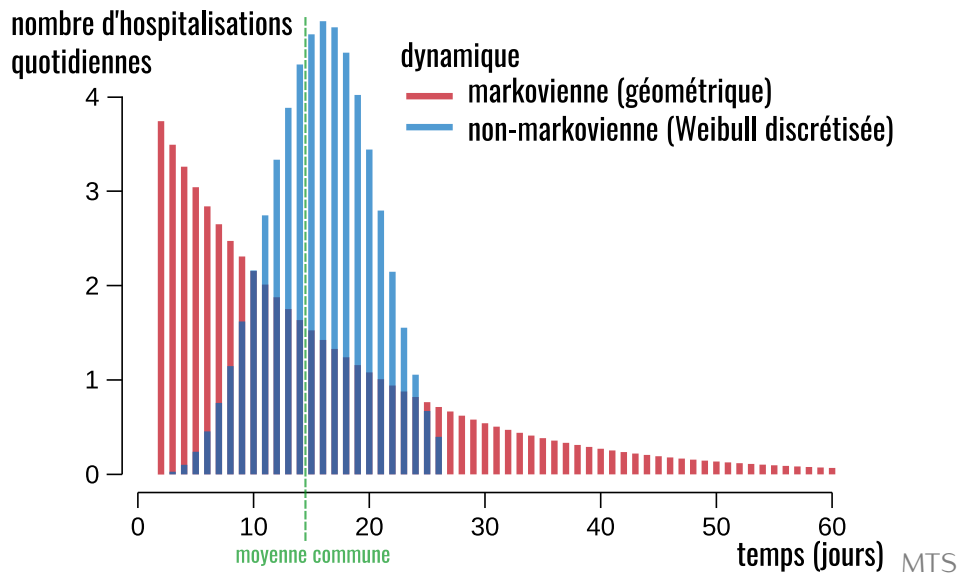


FIGURE 3.5 – **Comparaison illustrative des distributions de dates d’hospitalisations avec et sans propriété de Markov.**

Supposons 100 personnes infectées le même jour ( $t = 0$ ) par un virus dont l’infection entraîne une hospitalisation systématique. Les barres représentent la distribution du nombre d’admissions quotidiennes des individus composant cette cohorte, selon une dynamique markovienne (sans mémoire, ici une loi géométrique (équivalent discret de la loi exponentielle), en rouge), soit une distribution non-markovienne (Weibull discrétisée, en bleu). Les deux distributions présentent strictement la même moyenne, 14,5 jours – l’estimation du délai moyen entre la contamination et l’admission en soins critiques pour la première vague de COVID-19 en France (Sofonea et al., 2021c).

Afin de concilier parcimonie, intégration de cinétiques empiriques (de distribution quelconque<sup>\*</sup>), j’ai ainsi développé un modèle en temps discret (à pas quotidien, directement superposable aux données), dont le principe générique se formalise de la façon suivante :

$$\boxed{\mathcal{X}} \xrightarrow{\mathbf{u}} \boxed{\mathcal{Y}} \xrightarrow{\mathbf{v}} \dots \iff y_k(t+1) = (1 - v_{k-1})y_{k-1}(t) + \langle \mathbf{u} | \mathbf{x}(t) \rangle.$$

Ainsi, pour tout compartiment clinico-épidémiologique focal  $\mathcal{Y}$  duquel les individus sortent au taux  $\mathbf{v}$  (distribué selon le temps passé dans  $\mathcal{Y}$ ), et recevant un flux d’individus depuis un autre compartiment  $\mathcal{X}$  au taux  $\mathbf{u}$  (distribué selon le temps de résidence dans  $\mathcal{X}$ ), alors le nombre d’individus  $y_k$ , présents dans  $\mathcal{Y}$  depuis  $k$  jours<sup>†</sup>, est donné à la date  $t + 1$  par l’addition

\*. Une solution déjà utilisée dans la littérature pour échapper à la distribution exponentielle dans un système d’équations différentielles ordinaires s’appelle la méthode des stades (Cox and Miller, 1977), qui repose sur la division d’un compartiment en sous-compartiments (stades) chaînés (sans bifurcation). Au-delà des problèmes de paramétrisation (les stades n’ayant pas de réalité clinico-épidémiologique, leurs nombres et taux de transition intermédiaires ne sont pas identifiables), cette méthode ne peut générer tout au plus que des lois hypo-exponentielles.

†. Les temps de résidence négatifs ou nuls annulant naturellement la variable, i.e.  $\forall k \leq 0, y_k = x_k = 0$ .

- de la fraction d’individus déjà présents depuis  $k - 1$  jours la veille,  $y_{k-1}(t)$ , qui n’en sont pas sortis ce jour-là (avec probabilité  $1 - v_{k-1}^*$ ),
- et de la somme du nombre d’individus dans  $\mathcal{X}$  la veille pondérée par le vecteur  $\mathbf{u}$  selon leurs temps de résidence respectif (écrit sous forme de produit scalaire avec  $\mathbf{x}(t)$ ).

Le modèle effectif, détaillé dans sa première version dans l’article **Sofonea et al., 2021c** (voir tiré à part, p. 84), comprend deux niveaux d’ajustements supplémentaires.

Premièrement, la stratification en âge : il y a autant de compartiments  $\mathcal{Y}_i$  et de vecteurs  $\mathbf{u}_i$  que de classes d’âge  $i$ . Ceci est indispensable compte tenu du facteur de risque majeur de formes graves que représente l’âge <sup>†</sup> (Russell et al., 2020; Verity et al., 2020; O’Driscoll et al., 2020), mais aussi de l’hétérogénéité des couvertures vaccinales (base VAC-SI (Santé Publique France, 2021)).

Deuxièmement, les taux inter-compartimentaux sont dépendants du temps <sup>‡</sup>. Ceci permet d’accommoder la non stationnarité des

- processus gouvernant l’épidémie – en plus de la prévalence de l’infection (cf. *infra*) –,
  - les combinaisons d’interventions non pharmaceutiques (Ganser et al., 2023),
  - la transmissibilité différentielle des variants (Volz et al., 2021; Gaymard et al., 2021; Alizon et al., 2021; Sofonea et al., 2022b))
- ainsi que ceux gouvernant l’occupation des services hospitaliers :
  - l’amélioration des soins (Velly et al., 2020; Dennis et al., 2021),
  - la virulence différentielle des variants (Davies et al., 2021; Twohig et al., 2022; Nyberg et al., 2022), l’efficacité vaccinale (Liu et al., 2021)).

La procédure d’inférence repose sur une maximisation de vraisemblance classique couplée à une application du théorème de Wilks (Wilks, 1938), dont les détails sont fournis en annexe de **Sofonea et al., 2021c**. Ses principales limites relèvent de l’absence de spatialisation (bien qu’une stratification par département sera réalisée pour COVIDici à partir de 2021), de stochasticité et d’hétérogénéité de transmission.

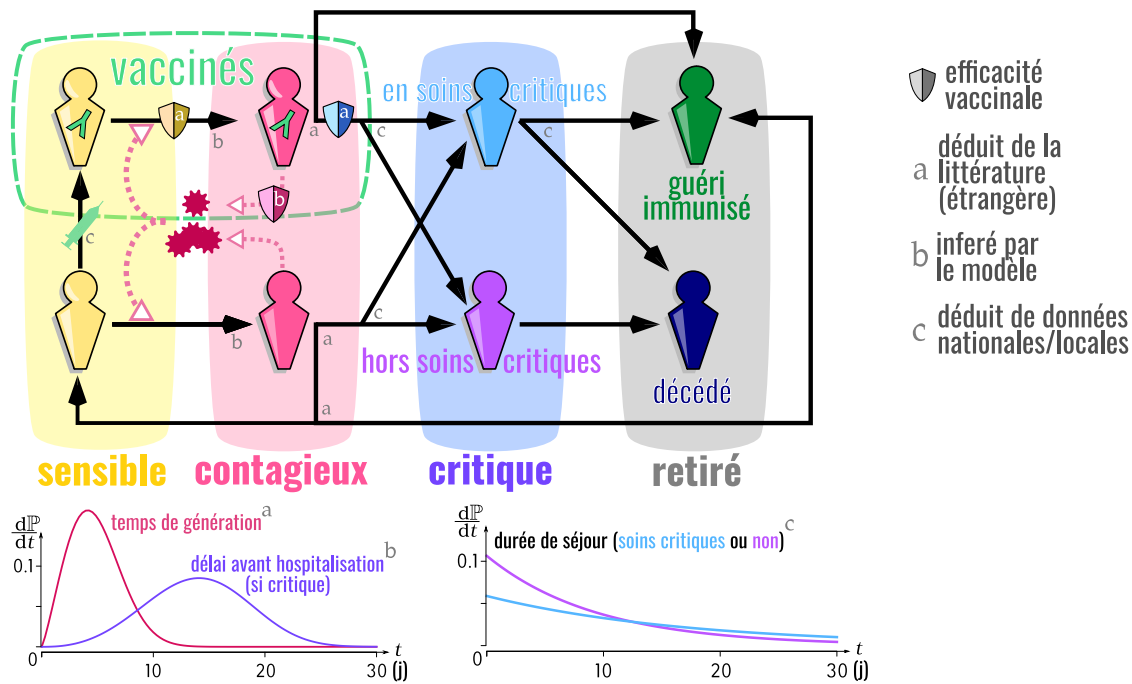
La structure finale du modèle (2021), intégrant notamment l’évolution virale et la vaccination, est présentée en fig. 3.6. Il sera dit central par la suite compte tenu de son implication récurrente dans les travaux ultérieurs dont les principaux résultats sont présentés ci-après.

---

\*. Où  $v_{k-1}$  est la composante du vecteur  $\mathbf{v}$  correspondant à la probabilité de sortir d’ $\mathcal{Y}$   $k - 1$  jours après y être entré. Les éléments de  $\mathbf{v}$  sont reliés aux données empiriques de la façon suivante :  $v_k := (F_v(k) - F_v(k-1)) / (1 - F_v(k-1))$ , où  $F_v$  est la fonction de répartition (quelconque) de la variable aléatoire correspondant au temps passé dans  $\mathcal{Y}$  (le raisonnement est analogue pour  $\mathbf{u}$ ). Dans la pratique (pour des raisons computationnelles), tous les éléments de  $\mathbf{v}$  sont nuls à partir d’un certain rang par troncature à droite du support de la variable aléatoire et renormalisation de la fonction de répartition (voir [annexes en ligne](#) de **Sofonea et al., 2021c**).

†. Mais il n’est pas le seul, voir en particulier **Semenzato et al., 2021** pour le sexe et les comorbidités. La flexibilité du modèle permet de les intégrer facilement, mais l’incertitude sur leur distribution jointe a représenté un frein important.

‡. Plus précisément, constants par fenêtre temporelle, dont les bornes étaient adaptatives.



**force d'infection relative à un individu sensible de classe d'âge  $i$  :**

$$\Lambda_i(t) = \left( 1 + \left( \mathcal{R}_0 q(t) c_i(t) \sum_j c_j(t) \sum_{\tau} \zeta_{\tau} \left( \pi_{j,\tau}^{\bar{V}}(t) + v_T \pi_{j,\tau}^V(t) \right) \right)^{-1} \right)^{-1}$$

Annotations for the equation:

- $\mathcal{R}_0$ : nombre de reproduction de base
- $q(t)$ : ratio de contact infectieux
- $c_i(t)$ : inférence interne
- $c_j(t)$ : inférence EpiEstim
- $\zeta_{\tau}$ : valeurs historiques supposées
- $\pi_{j,\tau}^{\bar{V}}(t)$ : fréquence dans la population des individus contagieux de la classe  $j$ , non vaccinés et infectés  $\tau$  jours plus tôt
- $v_T \pi_{j,\tau}^V(t)$ : efficacité vaccinale vis-à-vis de la transmission secondaire

Graphs for the equation:

- Graph 1:** Frequency of A0B9C0D1 vs. A0B9C1D0. Shows 'avantage de transmissibilité déduit des RT-qPCR de criblage' with a value of +96.5 [87.9 - 105]%. X-axis: 03 Dec, 23 Dec, 12 Jan, 01 Feb, 21 Feb.
- Graph 2:** Time  $t$  (semaines) from -3 to present. Legend: inférence interne (green), inférence EpiEstim (orange), valeurs historiques supposées (red).

MTS

FIGURE 3.6 – Vue d'ensemble du modèle épidémiologique central.

Chaque classe d'âge (une seule strate étant ici représentée) est divisée en compartiments (figurines) dépendants du statut infectieux, clinique et immunologique. Les individus sensibles (en jaune) peuvent être infectés s'ils sont exposés aux individus infectés de la communauté (en rose). Une fraction des personnes infectées développe une forme critique de COVID-19, ici définie comme nécessitant des soins critiques et/ou conduisant à un décès à l'hôpital. La vaccination (anticorps verts) confère trois protections (boucliers) : vis-à-vis de l'infection (jaune), de la transmission secondaire (rose) et vis-à-vis d'une forme critique (bleu).

Les densités de probabilités ( $\frac{dP}{dt} = f(t)$ ) correspondent aux cinétiques de contagiosité, de délai contamination - hospitalisation et de séjour (selon le service). Les lettres grisées indiquent l'origine des données. La force d'infection  $\Lambda_i(t)$  correspond à la probabilité pour un individu de la classe d'âge  $i$  d'être infecté le jour  $t$ . Elle capture la non-linéarité de la dynamique épidémique. Préférée à l'exponentielle obtenue par intégration (Diekmann et al., 2021) car elle converge plus lentement vers 1 (là où l'homogénéité de la population n'est plus envisageable). Sa valeur est dans la pratique très proche, pour les prévalences observées, de celle de leur tangente commune à l'origine, donnée par la quantité comprise par le deuxième couple de parenthèses. La démonstration complète est donnée en annexe de (Sofonea et al., 2021c).

### 3.1.2.2 Résultats fondamentaux

Le principal intérêt du modèle central est de combiner deux qualités habituellement exclusives :

- la **précision à court terme** apportée par les effets mémoire (dépendance de la contagiosité et de l'aggravation des symptômes à l'âge de l'infection) et plus largement l'intégration de distributions temporelles quelconques, qui est un des principaux avantages des modèles stochastiques,
- la **facilité d'implémentation** et le **coût computationnel minimal**, apanage des modèles déterministes (*a fortiori* en temps discret uniforme), permettant l'exécution d'un grand nombre de simulations en un temps réduit\* et donc l'exploration numérique intensive (p.ex. pour les algorithmes d'optimisation, les analyses de sensibilité) et la portabilité†.

Cette combinaison explique le regain d'attention portée à la modélisation épidémiologique en temps discret (Diekmann et al., 2021; Diekmann and Inaba, 2023), au moment où les modèles compartimentaux markoviens ne parvenaient pas à capturer le temps court de l'épidémie sur plusieurs indicateurs à la fois, comme le montre la fig. 3.7. Ce dernier constat a d'ailleurs motivé, au sein de l'équipe, le développement formel de modèles non markoviens en temps continu, au moyen d'équations aux dérivées partielles (Reyné et al., 2022a).

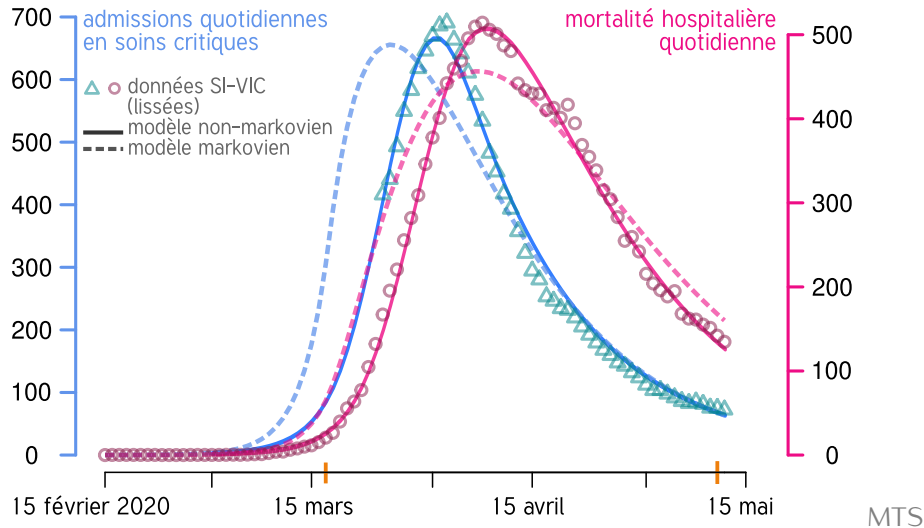


FIGURE 3.7 – **Effet mémoire & ajustement à la première vague hospitalière en France (printemps 2020).**

Meilleurs ajustements conjoints (par moindres carrés) aux données SI-VIC (Santé publique France, 2020a) quotidiennes lissées (moyenne glissante sur 7 jours des admissions en soins critiques et mortalité hospitalière COVID-19) du modèle central (traits pleins) et de son équivalent markovien en temps continu (pointillés).

\*. En local sur une machine ordinaire, et des sorties quasi-immédiates à chaque changement de paramètre par l'utilisateur sur l'application en ligne.

†. En particulier dans des situations techniquement dégradées.



Une fois ajusté sur l'épidémie en France métropolitaine (avril 2020), le modèle central a permis l'exploration théorique de scénarios alternatifs de contrôles. Au moyen de simulations contre-factuelles à date de confinement variable (et toutes choses égales par ailleurs), j'ai estimé que si le premier confinement avait été instauré une semaine plus tôt, resp. plus tard, **la mortalité hospitalière de la première vague aurait été minorée de 13 000 décès, resp. majorée de 46 000 décès** (quantiles à 2,5 %) (Sofonea et al., 2021c).

Face à une hypothétique reprise épidémique postérieure au premier déconfinement, il fallait envisager une stratégie plus soutenable qu'un nouveau confinement long et strict, ce qui nous a amené en mai 2020 à tester virtuellement de nouveaux patrons de réponse sanitaire et à en comparer les bilans prospectifs (sur le reste de l'année 2020) en terme de mortalité hospitalière et de potentiel relâchement (en durée, intensité et fraction de la population). J'ai ainsi montré qu'un contrôle différentiel selon la vulnérabilité ne présentait qu'un bénéfice marginal : à moins d'imposer des restrictions de contact plus importantes que celles du premier confinement aux personnes plus âgées pour compenser le relâchement du reste de la population, la mortalité cumulée était comparable à celle d'un contrôle uniforme sur toute la population (Sofonea et al., 2020b).

Poursuivant l'idée introduite par Ferguson et al., 2020 d'une alternance entre confinements et périodes de relâchement indexée sur un indicateur hospitalier, j'ai montré, en cherchant à minimiser le bilan du confinement adaptatif\*, que la proportion de temps hors confinement dépend très peu du seuil de déclenchement de ce dernier, lequel joue en revanche fortement sur la mortalité, comme l'illustre la fig. 3.8. Autrement dit, **ce résultat suggère que retarder une mesure de contrôle face à une reprise épidémique alourdit l'impact hospitalier sans pour autant diminuer la durée totale des restrictions sur le temps long.**

J'ai par ailleurs étudié l'intérêt d'alternances périodiques†, non dictées par un indicateur. J'ai notamment montré que la mortalité était d'autant plus réduite que la fréquence d'alternance était élevée (et qu'en particulier un jour et demi de relâchement modéré par semaine pendant un confinement était envisageable), mais que cette stratégie était moins prometteuse qu'un confinement adaptatif optimisé (Sofonea et al., 2020b).

---

\*. Un temps appelé stratégie du *stop and go* dans les médias (Berrod, 2020).

†. Le confinement le week-end sera instauré un an plus tard à Dunkerque et à Nice (Thibert, 2021).

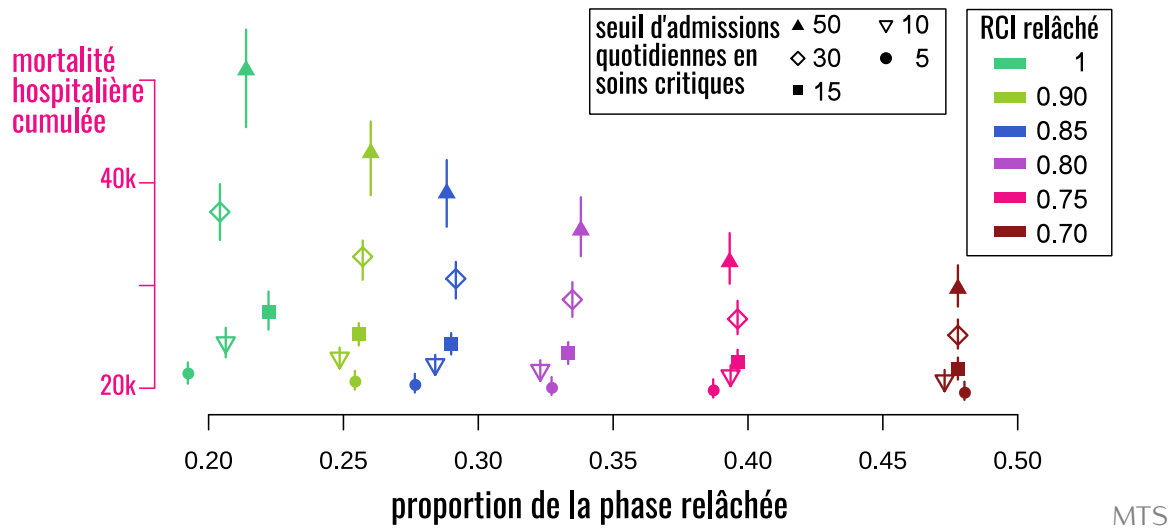


FIGURE 3.8 – **Exploration numérique du confinement adaptatif.**

La stratégie de confinement adaptatif, évoquée par Ferguson et al., 2020 le 16/03/2020, consiste à déclencher un confinement dès qu'un indicateur dépasse un certain seuil  $s$ , le relâcher dès qu'il repasse sous un second seuil  $s'$ , en maintenant éventuellement certaines mesures moins contraignantes en place entre chaque confinement (le principe est extrapolable à toute mesure de prévention comportementale). Ici, les deux seuils correspondent à un nombre quotidien d'admissions COVID en soins critiques, au niveau national, et sont supposés identiques,  $s = s'$ . Le deuxième paramètre est mesuré en ratio de contact infectieux (RCI) pendant les périodes de relâchement. Le RCI est donné par le rapport entre le taux moyen de contact potentiellement infectieux d'un individu à un instant donné après le 17 mars 2020, relativement à la valeur de ce même taux moyen avant cette date (un RCI de 1 correspond donc à un retour complet aux habitudes de contact pré-pandémiques). Ici, chaque figuré correspond à la simulation d'un confinement adaptatif, hypothétiquement mis en œuvre à partir du 12 mai 2020 (début du premier déconfinement), et paramétré par un couple  $(s, \text{RCI})$  de valeurs respectivement signifiées par la forme et la couleur du figuré (voir légende). L'abscisse du figuré indique la proportion médiane du temps passé dans la phase de relâchement de chaque cycle (p.ex. : un RCI relâché élevé engendre une forte reprise épidémique, laquelle déclenche rapidement un nouveau confinement, limitant la durée de la phase relâchée). L'ordonnée représente la médiane du nombre de décès hospitaliers cumulée jusqu'à la fin de l'année 2020 (sur la base des informations disponibles mi-mai 2020). Les barres verticales couvrent 95 % des valeurs des simulations. (Résultats produits en mai 2020, non (pré-)publiés).

La portée de ces résultats reste cependant limitée compte tenu des hypothèses simplificatrices qui les sous-tendent, notamment ici l'absence de spatialisation (Karatayev et al., 2020), de *fatigue pandémique*\* (Delussu et al., 2022) et de considérations socio-économiques (Pangallo et al., 2022).

\*. Diminution de l'adhésion/compliance de la population aux mesures sanitaires.

### 3.1.2.3 Résultats appliqués

L'ajustement du modèle central aux séries temporelles hospitalières nationales a permis d'estimer les principaux paramètres épidémiologiques au début du mois d'avril 2020. Les estimations centrales ont faiblement fluctué les semaines suivantes tandis que l'incertitude s'est réduite. À la sortie du premier confinement, nous estimions ainsi qu'en France métropolitaine, le **nombre de reproduction de base ( $\mathcal{R}_0$ ) était égal à 3,0 [2,6;3,4]** (intervalle de vraisemblance à 95 %), qu'il avait été **réduit de 76 [73;79] % de sa valeur au cours du confinement**, et que 2,4 [2,3;2,5] % de la population française avait été infectée (Sofonea et al., 2021c). Ces estimations se sont trouvées être proches (bien que tous les intervalles ne se superposent pas) de celles des autres études contemporaines (Salje et al., 2020; Roques et al., 2020; Hoertel et al., 2020; Forien et al., 2020). La transposition du modèle central aux données irlandaises\* a permis d'estimer la réduction du nombre de reproduction à 64 [44;75] % au cours du premier confinement sur cette île (Kamiya et al., 2023).

Les résultats du modèle central ont aussi permis de calibrer deux approches méthodologiquement complémentaires. La première, une modélisation stochastique avec super-propagateurs (l'hétérogénéité individuelle de transmission étant alignée sur celle du SARS-CoV (Lloyd-Smith et al., 2005)), a ainsi estimé le début de la vague épidémique en France (date d'arrivée du premier cas ayant initié une chaîne de transmission ininterrompue) au 19 janvier 2020, intervalle de confiance à 95 % entre 29 décembre 2019 et le 29 janvier 2020 (Beneteau et al., 2021a). Ce modèle suggérait aussi qu'il aurait théoriquement fallu que le confinement dure 7 mois pour éradiquer (avec une probabilité de 97,5 %) le SARS-CoV-2. La seconde approche est le logiciel Epidemap, développé par Olivier Thomine en 2020, capable de simuler les déplacements quotidiens des 66 millions de français métropolitains à l'échelle des bâtiments eux-mêmes, par combinaison de noyaux de mobilité aux données spatialisées d'OpenStreetMap. L'ajout d'une sur-couche épidémiologique stochastique dans le logiciel, calibrée par le modèle central, a permis de mettre en évidence les propriétés émergentes comme la dépendance concave du délai avant l'arrivée du virus dans une commune à la distance de celle-ci à l'origine spatiale de l'épidémie, la modulation de cette dépendance à la densité de la commune ou bien la corrélation positive entre cette dernière et l'espérance du taux d'attaque (le résultat classique (Miller, 2012) de la taille finale de l'épidémie en champ moyen † n'étant vérifié que pour les grandes agglomérations) (Thomine et al., 2021, voir tiré à part p. 102).

Au printemps 2021, le modèle central a intégré le déploiement de la vaccination et l'arrivée des variants Alpha et Delta. Si les efficacités vaccinales vis-à-vis de l'infection et des formes graves ont été rapidement estimées, y compris contre ces variants, (Puranik et al., 2021; Public Health

---

\*. En collaboration avec les collègues de l'*University of Galway* (République d'Irlande) et en particulier Dr. Tsukushi Kamiya.

†. Hypothèse de travail simplificatrice considérant que chaque personne peut interagir avec n'importe quel autre individu de la population avec la même probabilité d'occurrence et la même intensité. Tout se passe comme si toute la population d'étude était concentrée dans un même point. Malgré son irréalisme évident, le biais introduit par cette hypothèse reste limité pour un certain nombre de questions (Trapman et al., 2016), tandis que la paramétrisation de son alternative peut s'avérer difficile voire impossible.

England, 2021), l'éventuelle réduction de la contagiosité conférée par la vaccination ne faisait l'objet que d'estimations très préliminaires à ce moment-là. Le modèle central a ainsi fourni une des premières quantifications par modélisation de cette protection, à savoir une **réduction de 43 [32;55] % de la transmission secondaire** (conditionnellement à l'infection; en moyenne sur les 3 premiers mois post-première dose), compatible avec les résultats ultérieurs de la littérature (Prunas et al., 2022), y compris à l'ère d'Omicron (Tan et al., 2023). Sur le même sujet, nous avons simulé la dynamique épidémique en France de janvier à août 2021 en annulant l'effet des vaccins (tout en maintenant l'arrivée des variants Alpha, Delta, et d'éventuelles réponses sanitaires permettant d'éviter la saturation hospitalière). La confrontation aux données SI-VIC du bilan contrefactuel ainsi généré, et présenté en fig. 3.9, a permis de fournir une première quantification de l'impact en vie réelle de la vaccination en France métropolitaine (sur les neuf premiers mois de son déploiement). Nous avons ainsi estimé que la vaccination a prévenu au moins 26 000 séjours en soins critiques et plus de 36 000 décès hospitaliers (quantiles à 2,5 %) (Sofonea et al., 2021a). Ce deuxième point équivaut à **11 [8;14] (intervalle de confiance à 95 %) décès évités pour 10 000 doses de vaccins**, une valeur cinq fois inférieure aux estimations de Watson et al. 2022 pour la France, qui s'explique par l'absence de contrôle de l'épidémie dans leur modèle, ce qui attribue aux vaccins une réduction de la mortalité qui, en vie réelle, aurait été en fait due à des interventions non pharmaceutiques.

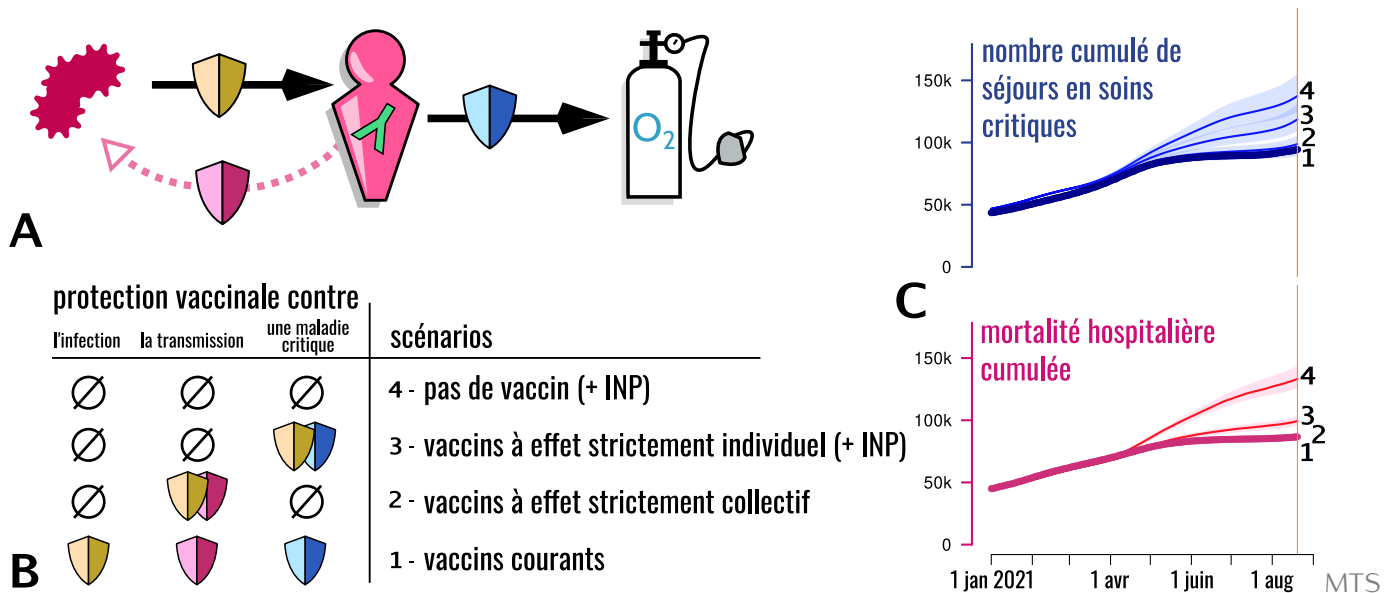


FIGURE 3.9 – Quantification par contrefactuels de l'impact hospitalier en vie réelle de la vaccination anti-SARS-CoV-2 (France, janvier 2021 - août 2021).

**A** Aperçu schématisé des trois protections (partielles) conférées par la vaccination anti-SARS-CoV-2 : contre l'infection (bouclier jaune) si exposé au SARS-CoV-2 (particules grenat), la transmission secondaire si infecté (bouclier rose), et des complications nécessitant des soins critiques (bouclier bleu). **B** Les scénarios contrefactuels incluent toutes les composantes prophylactiques du vaccin (1), aucune (4) ou bien ne présentent qu'un effet collectif (2) ou individuel (3). L'effet de la composante anti-infectieuse étant à la fois individuel et collectif, la réduction correspondante doit être transférée aux autres composantes dans les scénarios 2 et 3 pour distinguer les bénéfices directs et indirects du vaccin. Les scénarios 3 et 4 entraînant une augmentation massive des infections, des IPN (interventions non pharmaceutiques) sont automatiquement implémentées dans ces scénarios afin de reproduire un évitement réaliste de la saturation nationale des services de soins critiques. (NB : ce choix diverge de celui opéré par Watson et al., 2022). **C** Séjours en soins critiques et mortalité hospitalière étiquetés COVID en France entre le 1er janvier et le 20 août 2021 (la campagne de vaccination ayant débuté le 27 décembre 2020). Les courbes épaisses correspondent aux comptes cumulés effectifs, sur lesquels est ajusté le scénario n°1. Les numéros de scénarios renvoient à ceux du panneau B. Les courbes pleines et rubans associés représentent respectivement la médiane et l'étendue de 95 % des simulations. La quantification de l'impact de la vaccination en vie réelle est donnée par la différence entre les scénarios n°1 et 4.

L'implémentation du modèle central s'est traduite par le développement de deux applications phares de l'équipe, COVIDSIM-FR (Sofonea et al., 2020a) et COVIDici (Boennec et al., 2021b), librement accessibles en ligne respectivement le 06/04/2020 et le 02/02/2021 \*. Ces deux applications ont principalement été utilisées pour anticiper l'occupation des services de soins critiques, respectivement au niveau national et infra-national (régions et départements). COVIDSIM-FR a notamment permis de réaliser des projections régulières (en moyenne 2,5 scénarios par période), à des moments charnières de l'épidémie (rebonds, pics, déconfinements (Sofonea et al.,

\*. Elles ont connu plusieurs mises à jours et améliorations par la suite.

2021b), nouveaux variants (Alizon et al., 2021; Sofonea et al., 2022b)). La rétrospective de la première année de projections (septembre 2020 - septembre 2021) a montré qu'en moyenne, sur les huit périodes concernées, l'anticipation par le meilleur scénario (lequel n'a été le plus optimiste qu'à 3 reprises) était **correcte jusqu'à 38 jours en avance** (Sofonea and Alizon, 2021, cf. tiré à part p. 98). Six mois plus tard, COVIDSIM-FR projetait avec justesse que la suspension du passe vaccinal\* et la fin du port obligatoire du masque en intérieur† n'allait pas engendrer de rebond hospitalier comparable aux précédentes vagues (Sofonea et al., 2022b, cf. tiré à part p. 116).

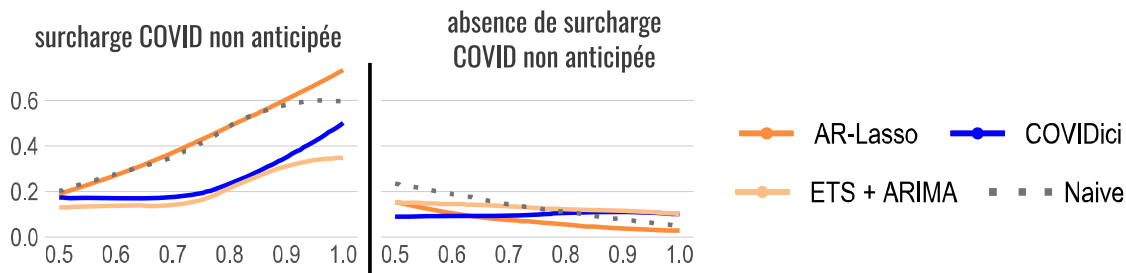
En parallèle, les performances de COVIDici sur l'année 2021 ont été comparées à des méthodes de *machine learning* appliquées aux séries temporelles. Comme le montre la fig. 3.10, **aucune des méthodes alternatives testées ne fait uniformément mieux que COVIDici pour un horizon de 4 semaines** (c'est-à-dire à la fois en terme d'anticipation correcte de surcharge et d'anticipation correcte d'absence de surcharge des soins critiques, et en ce sens sur l'ensemble des 13 régions métropolitaines) (Massey et al., 2023). Faisant jeu égal avec les approches phénoménologiques sur le temps long, mais bénéficiant en plus d'une paramétrisation identifiable *a priori* (laquelle permet d'inclure pro-activement les changements induits par l'arrivée d'un nouveau variant ou le relâchement programmé d'une mesure sanitaire), l'expérience de COVIDSIM-FR et de COVIDici suggère que les modèles mécanistiques peuvent avec profit trouver leur place dans l'arsenal d'aide à la décision sanitaire, qu'il s'agisse de mesures en population ou de gestion hospitalière.

---

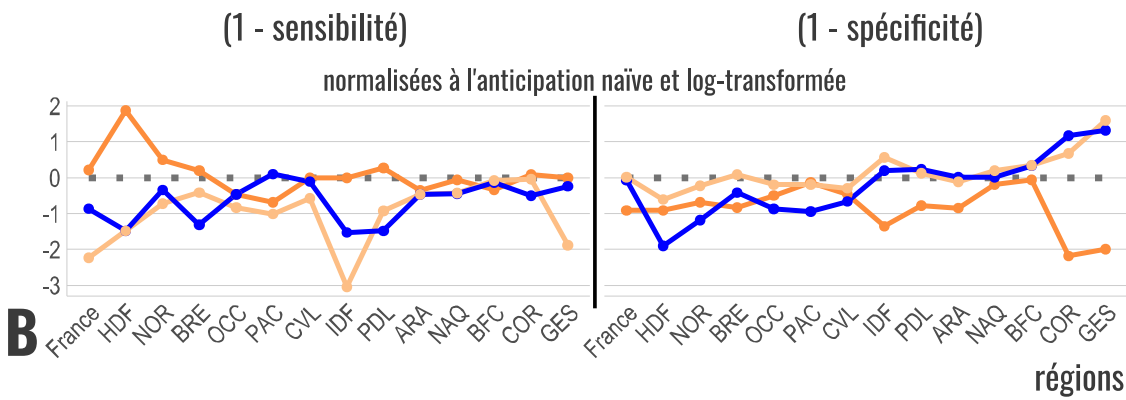
\*. À l'exception des établissements de santé, maisons de retraite, et établissements accueillant des personnes fragiles ou en situation de handicap.

†. À l'exception de l'ensemble des transports en intérieur et dans les lieux de santé ou de soins.

### fraction spatio-temporelle des projections erronées de la situation des services de soins critiques (à 4 semaines)



**A** ratio du pic local de la première vague (avril 2020)



régions A. Massey

**FIGURE 3.10 – Comparaison inter-méthodes de l'anticipation de la tension en soins critiques.**

Les projections régionales à 4 semaines de l'occupation des lits de soins critiques par des patients COVID produites par COVIDici (application automatisée et administrativement stratifiée du modèle central; courbes bleues) entre le 30/01/2021 et le 02/12/2021, sont comparées à celles générées par trois approches phénoménologiques à partir des mêmes données. Il s'agit de *naive* (pointillés gris) : prédiction constante égale à la dernière valeur connue, d'*ETS+ARIMA* (orange clair) : une moyenne entre un lissage exponentiel (*ETS*) et une autorégression à moyenne mobile intégrée (*ARIMA*) et d'*AR-Lasso* (orange foncé) : un modèle de *machine learning* associant une auto-régression et une pénalisation lasso. *Naive* et *ETS+ARIMA* ont été implémentées avec le paquetage R *fable* (O'Hara-Wild et al., 2020), *AR-Lasso* avec le paquetage *caretForecast* (Akay, 2022). **A**) Fraction agrégée en temps (du 30/01/2021 au 02/12/2021) et en espace (13 régions de France métropolitaine) des projections n'ayant pas anticipé une surcharge COVID en soins critiques (à gauche), ou ayant anticipé à tort cette dernière (à droite), pour chacune des 4 approches, en fonction du seuil arbitraire de surcharge (rapporté à la valeur locale du pic de la première vague). **B**) Agrégation temporelle des complémentaires de la sensibilité (i.e. fraction d'anticipation de surcharge à tort, à gauche) et de la spécificité (i.e. fraction d'anticipation de sous-charge à tort, à droite), normalisées par rapport à la projection *naive* et log-transformée, pour la France métropolitaine et l'ensemble de ses 13 régions.

ARA : Auvergne-Rhône-Alpes, BFC : Bourgogne-Franche-Comté, BRE : Bretagne, CVL : Centre-Val de Loire, COR : Corse, GES : Grand Est, HDF : Hauts-de-France, IDF : Île-de-France, NOR : Normandie, NAQ : Nouvelle-Aquitaine, OCC : Occitanie, PDL : Pays de la Loire, PAC : Provence-Alpes-Côte d'Azur, France : France métropolitaine.

### 3.1.3 Contributions additionnelles

Je résume ci-après des contributions à la recherche sur le SARS-CoV-2 qui ne découlent pas du modèle évoqué dans les pages précédentes.

#### 3.1.3.1 Santé publique

Dans les premières semaines de la pandémie, nous avons cherché à déterminer des stratégies minimisant à la fois le risque de saturation hospitalière et le poids des restrictions. Nous avons appliqué la théorie du contrôle optimal pour déterminer l'intensité de distanciation sociale au cours du temps qui minimise à la fois la surmortalité hospitalière induite par le défaut de soins et le coût des mesures (Djidjou-Demasse et al., 2020a). Nous avons en particulier montré que la stratégie mathématiquement optimale consiste à réduire immédiatement les interactions de la fraction la plus âgée de la population, avant de permettre un relâchement progressif – la tranche d'âge et la cinétique étant dépendant dans le détail de la démographie du pays considéré (les travaux ont porté sur la France, le Vietnam et le Burkina-Faso) (Richard et al., 2021). À la sortie du premier confinement, nous avons mis en évidence les bénéfices du port du masque et de la sectorisation COVID dans les EHPAD sur la base du dépistage régulier au sein de 12 centres de l'Hérault au cours de la première vague (Reyné et al., 2021).

#### 3.1.3.2 Virologie

La pandémie de SARS-CoV-2 a entraîné une utilisation massive sans précédent des tests de dépistage, essentiellement à usage diagnostique mais aussi clinique (estimation indirecte de la charge virale) et d'épidémiologie-surveillance (criblage de variants). Avec un consortium national de virologues, nous avons montré, en suivant l'idée d'Hay et al. (2021), que les nombres de cycles d'amplification (valeurs Ct produites par les RT-qPCR<sup>\*</sup>) peuvent être utilisées comme information complémentaire pour améliorer les prédictions à court terme pour la surveillance des épidémies, à condition de disposer de grands jeux de données (près de 800 000 dépistages positifs dans notre cas) et de corriger pour des cofacteurs tels que l'âge (Alizon et al., 2022). L'analyse des données de dépistage du Laboratoire Cerba<sup>†</sup>, complétées par celles de l'AP-HP et du CHU de Montpellier<sup>‡</sup>, nous a aussi permis de montrer que les valeurs Ct des variants du SARS-CoV-2 apparus à partir de l'automne 2020 étaient inférieures à celles atteintes par la souche historique (Roquebert et al., 2021a), que les durées d'infection pouvaient pour partie le différentiel de transmissibilité et que ce dernier varie avec la démographie de la population considérée (Elie et al., 2022a). Enfin, nous avons attiré l'attention sur le fait que, compte tenu de la nature d'ARN simple brin à polarité positive du matériel génétique du SARS-CoV-2, il

---

\*. Réaction en chaîne par polymérase, quantitative et avec transcriptase inverse.

†. Représentant près d'un tiers de la couverture nationale.

‡. *Analysing the French COVID-19 Epidemic Using a National SARS-CoV-2 RT-PCR Database (CovPCR)* : NCT04738331 (dir. : Pr Vincent Foulongne, CHU Montpellier).



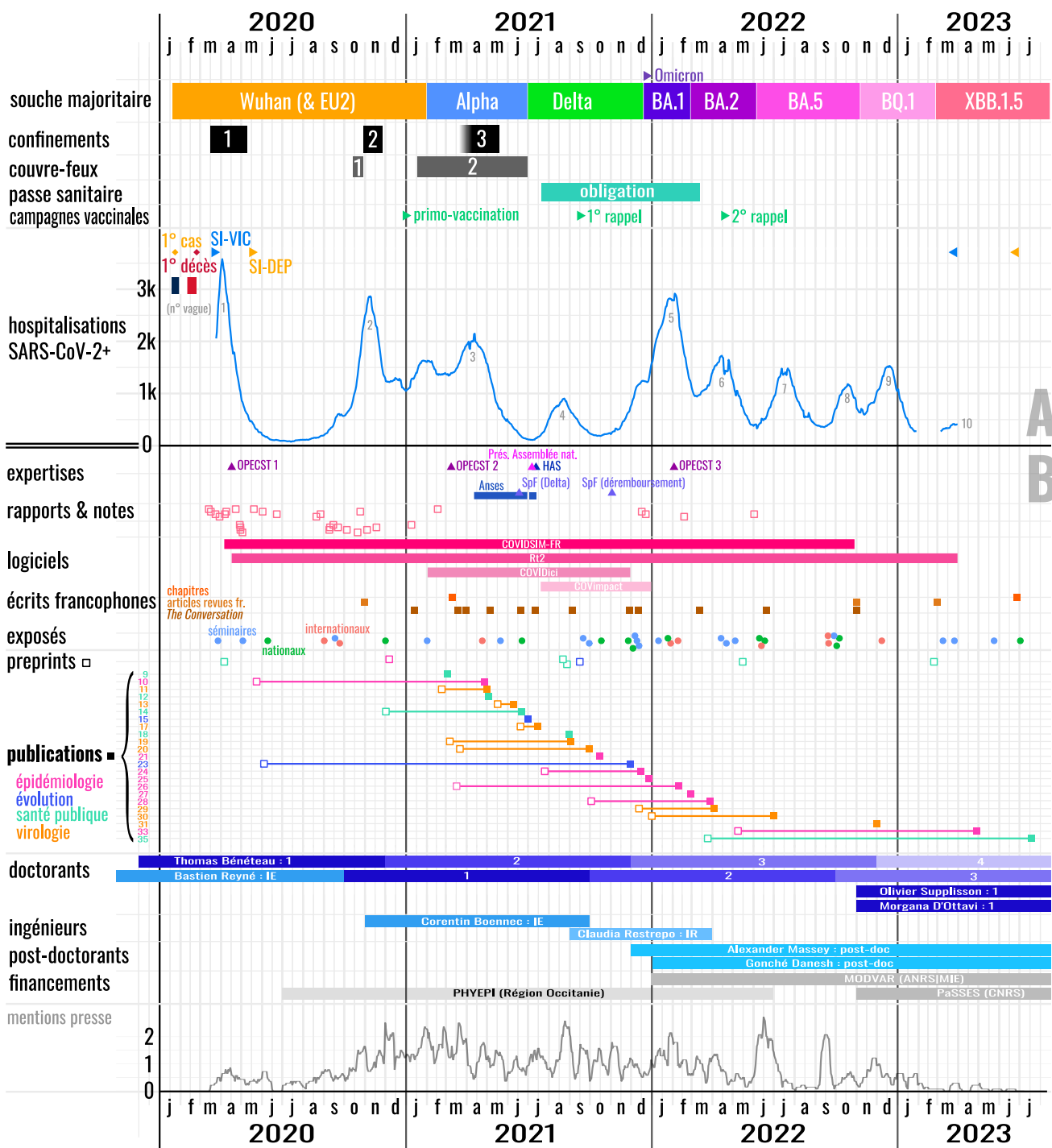
n'est pas possible distinguer l'ARN répliatif de l'ARN transcriptionnel, ce qui empêche d'interpréter directement les variations de Ct en terme de charge virale (Michalakakis et al., 2021).

### 3.1.3.3 Évolution

Le rôle joué par les variants du SARS-CoV-2 qualifiés de préoccupants (VOC) par l'OMS dans l'épidémiologie du COVID-19 moins d'un an après le début de la pandémie rappelle l'importance de la biologie évolutive dans la compréhension, la surveillance et l'anticipation de la dynamique virale. D'une part, nous avons utilisé les méthodes de phylogénie et de phylodynamique pour extraire le signal épidémiologique contenu dans les séquences génomiques de SARS-CoV-2, pour mesures des nombres de reproduction (Danesh et al., 2021) ou mettre en évidence des lignées virales à surveiller (Verdume et al., 2021). D'autre part, nous avons décrit la succession des principaux variants (Alpha, Beta, Gamma, Delta, Omicron BA.1 et BA.2) en France au moyen d'approches issues de la génétique des populations (Haim-Boukobza et al., 2021; Roquebert et al., 2021b; Alizon et al., 2021; Roquebert et al., 2021a; Sofonea et al., 2022b). En particulier, nous étés les premiers en France à estimer, en juin 2021, l'avantage de transmissibilité du variant Delta vis-à-vis du variant Alpha (à savoir +79 [52;110] % – intervalle de confiance à 95%), alors ultra-majoritaire (cette étude avait fait l'objet d'une note transmise à la Direction Générale de la Santé) (Alizon et al., 2021, cf. tiré à part p. 92). Les interrogations autour de la sévérité des nouveaux variants ont par ailleurs ravivé l'intuition collective de l'évolution des pathogènes vers l'avirulence. Nous avons alors rédigé une revue rappelant la littérature à ce sujet, mis en avant les pistes expliquant pourquoi la virulence du SARS-CoV-2 n'a pas diminué et exploré les scénarios d'évolution à plus long terme (Alizon and Sofonea, 2021).

### 3.1.3.4 Synthèse chronologique

La fig. 3.11 résume chronologiquement l'ensemble des activités scientifiques et péri-scientifiques relatives à la pandémie mentionnées dans les trois premiers chapitres de ce document.



MTS

FIGURE 3.11 – Synthèse chronologique des activités scientifiques et péri-scientifiques relatives à la pandémie de COVID-19, replacée dans le contexte sanitaire français (janvier 2020 - juillet 2023).

**A)** Dates et périodes remarquables de l'épidémie de COVID-19 en France (métropolitaine). De haut en bas : souche virale majoritaire dans les séquences françaises (données GISAID et EMERGEN) : ancestrale ou proche (EU2), variant préoccupant (VOC) au sens de l'OMS (Alpha, Delta) puis lignées Omicron (nomenclature Pango); principales interventions sanitaires nationales; dates des premiers cas et décès français, de fonctionnement des bases de données SI-VIC (hospitalières) et SI-DEP (dépiages) (Santé publique France, 2020a,b); hospitalisations quotidiennes nationales de patients SARS-CoV-2+ (moyenne glissante sur 7 jours) [NB : données non fiables en février 2023 et arrêt de la base un mois plus tard].

**B)** Activités collectives et personnelles sur le sujet. De haut en bas : expertises auprès des autorités et agences nationales (cf. p. 9); rapports et logiciels produits avec l'équipe ETE; vulgarisation écrite; travaux (stratifiés par discipline dominante (couleur) du manuscrit – numéros introduits p. 14); ressources humaines; mentions ('Mircea Sofonea') dans la presse lissées sur 2 semaines (données Europresse).

### 3.1.4 Conclusion

La pandémie a placé la recherche scientifique, comme bien d'autres domaines, dans une posture inédite : course à l'analyse en temps réel de données (parfois très) préliminaires, flux massif de (pré-)publications, fortes attentes de la société. Outil de l'incertitude et du manque de données, la modélisation s'est révélée être une plate-forme propice aux interactions entre les nombreuses spécialités scientifiques et médicales intéressées par les maladies infectieuses, et a fourni de précieux éléments pour la compréhension, la surveillance, le contrôle du COVID-19 et l'organisation hospitalière afférente (Gayat and Raux, 2022). Nous retiendrons qu'une approche mécanistique parcimonieuse et non markovienne peut être particulièrement utile pour répondre aux premières questions de santé publique posées par une maladie infectieuse émergente, et notamment en ce qui concerne l'anticipation des besoins hospitaliers, mais que ses faiblesses constitutives méritent d'être complétée par des approches phénoménologiques (Paireau et al., 2022, Massey et al., 2023) et individus-centrées (Karatayev et al., 2020, Thomine et al., 2021, Pangallo et al., 2022).

De véritables bilans critiques pluridisciplinaires restent encore à dresser, en intégrant d'autres métriques de santé (COVID long (Davis et al., 2023), sur-risques cardiovasculaires post-aigus (Xie et al., 2022), conséquences indirectes (Hillis et al., 2021)) que celles qui ont focalisé l'attention collective jusqu'à présent (lesquelles restent elles-mêmes à corriger (Wang et al., 2022; Msemburi et al., 2023)), ainsi que la contribution des sciences humaines et sociales. Cette démarche est salutaire pour l'abord de deux défis majeurs qui nous font face en matière de maladies infectieuses. Le premier concerne l'évolution de notre surveillance et des moyens collectifs consentis au contrôle des viroses respiratoires (Flahault et al., 2023), notamment par le levier de la qualité de l'air intérieur (Lewis, 2023), dont l'amélioration intéressera aussi l'exposome non infectieux (Lewis et al., 2023). Le second relève de la préparation scientifique, sanitaire, opérationnelle mais aussi culturelle aux futures pandémies et risques épidémiques émergents, lesquels se retrouvent majorés par le changement climatique (Mora et al., 2022). Mes futurs projets de recherche, présentés dans les sections suivantes, s'inscrivent dans ces deux perspectives.

Les deux sections suivantes relatives aux perspectives de recherche sont rédigées en anglais car elles reprennent deux projets soumis en 2023, dans la continuité des travaux réalisés au cours de la pandémie de SARS-CoV-2. Ils portent respectivement sur les virus respiratoires (court terme) et sur le *Pathogène X* (long terme).

## 3.2 Perspectives à court terme

### Surveillance, modelling & Control of Respiratory viruses (SCoRe)

#### 3.2.1 Background

From the first probabilistic case estimates made less than three weeks after the first atypical pneumonia cases detected in Wuhan in late December 2019 (Imai et al., 2020), to simulations seeking to optimise non-pharmaceutical interventions (Sonabend et al., 2021), epidemiological modelling has been at the forefront of scientific research and health policy-making over the last 3 years (Adam, 2020). The diversity of the determinants of the SARS-CoV-2 epidemiology, including spatial (Karatayev et al., 2020; Gaudart et al., 2021) and individual heterogeneities (Yang et al., 2021), airborne transmission (Peng et al., 2022), inoculum size (Lind et al., 2023), viral evolution (Markov et al., 2023), vaccination (UKHSA, 2023) and waning immunity (Goldberg et al., 2022), has further strengthened the need for quantitative and transdisciplinary analyses, the quality and relevance of which cannot be achieved through their mathematical, physical or computational features alone but essentially rely on the biological knowledge of the disease and public health context (Zelner et al., 2021; Anderson et al., 2021; Sofonea et al., 2022a).

The ambition of this project is motivated by the following observations.

- Before the pandemic and even more today, respiratory infections are the first communicable, hence controllable, cause of deaths worldwide (WHO, 2020; Vos et al., 2020). It especially affects the most fragile people and is often underdocumented and collectively underconsidered (Ferkol and Schraufnagel, 2014; Villanueva et al., 2022).
- The circulation of common respiratory pathogens such as influenza viruses and respiratory syncytial virus (RSV), so far suppressed by non pharmaceutical interventions (NPI) (Bonacina et al., 2023), is now fully resumed (Équipes de surveillance de la grippe (Santé publique France), 2022; Santé publique France, 2023). This amplifies morbi-mortality and health-service demand, especially regarding coinfection (Stowe et al., 2021; Swets et al., 2022) paediatric critical care (Roucaute, 2023).
- The eradication of respiratory viruses or even long-term herd immunity against them are out of reach with current means. In the case of SARS-CoV-2, the high contagiousness reached by the Omicron group (Sofonea et al., 2022b), the partial and waning effectiveness of both existing vaccines (Tartof et al., 2022; UKHSA, 2023) and post-infection immunity (Pulliam et al., 2022; Goldberg et al., 2022), and the return to pre-pandemic habits (Delussu

et al., 2022) expose the world to the emergence of immune escape and/or more virulent variants challenging the vaccine-based long-term control strategy (Markov et al., 2022; Sasaki et al., 2022; Callaway, 2023).

- The continuous decrease in surveillance effort jeopardizes our ability to early detect new variant emergence, importation or epidemic rebound in the community hence from triggering early and more sustainable public health responses (Bhatia et al., 2023).

In these circumstances, it appears essential and timely to reconsider the SARS-CoV-2 variant-specific surveillance while jointly monitoring the spread of the other major respiratory viruses and, if necessary, quickly trigger scalable and appropriate scientific and health responses (Flahault et al., 2023; Bhatia et al., 2023). Unfortunately, the epidemiological frameworks that were instrumental at the beginning of a pandemic have become inappropriate with the increasing heterogeneity in host susceptibility (due to post-infection and/or vaccine immunity), virus phenotypic diversity and now require the integration of multiple infections (Andreasen et al., 1997; Calvez et al., 2005).

The SCoRe project will be based on three pillars in which I have previous experience : virological data analysis (Danesh et al., 2021; Alizon et al., 2021; Roquebert et al., 2021a), epidemiological modelling (Sofonea et al., 2021c; Richard et al., 2021; Thomine et al., 2021) -- especially on multi-pathogen dynamics (Sofonea et al., 2015, 2017) -- and public health decision support (Le Cann et al., 2021; Sofonea and Alizon, 2021; Sofonea et al., 2022b). The project will aim (1) to monitor the SARS-CoV-2 diversity and incidence along with the two major co-circulating respiratory virus : seasonal influenza and RSV (2) to finely quantify the spatial component of their spreading and (3) to develop a methodological toolbox capable of providing rapid and robust projections to inform public health decision-making. More precisely, the synthetic framework will provide, through a common pipeline for the SARS-CoV-2/influenza/ RSV trio, inference of epidemiological parameters, month-scale projections (Massey et al., 2023), and counterfactual analysis for policymaking support. Further, I outline the three tasks composing the project while corresponding material, methods and collaborations are detailed the next sections.

### 3.2.2 Task #1 : monitoring respiratory infection diversity

The massive circulation of the SARS-CoV-2, a single-stranded RNA virus subject to strong selection pressures (NPI, post-infectious and vaccine immunity), has been fuelling an uninterrupted viral evolution (Markov et al., 2023). The phenotypic mutants called variants (and sub-variants) are associated with increased contagiousness and/or immune escape (Volz, 2023). Furthermore, infections caused by some of these variants appear to be more virulent (Twohig et al., 2022), which raises both clinical and evolutionary questions (Alizon and Sofonea, 2021, Markov et al., 2022). Therefore, their early detection and epidemiological characterisation, especially the estimation of their relative contagiousness and virulence, is a crucial public health issue.

In this task, the surveillance of SARS-CoV-2 as well as that of influenza A and B and RSV will be ensured by analysing the routinely updated databases of RT-qPCR and genomic sequences data collected by the EMERGEN consortium and by Cerba. Analyses will include Bayesian hierarchical modelling, population genetics and phylodynamics (Bouckaert et al., 2019), completed by spatial statistics (Dellicour et al., 2020) and approximate Bayesian computing (Saulnier et al., 2017; Danesh et al., 2023). Meanwhile, historical RVI data will be gathered from the Santé publique France database and compared to data prior to the past years incidence series, before and during the COVID-19 pandemic.

### 3.2.3 Task #2 : understanding and characterising spatial spread

The majority of epidemiological models used to guide public health decisions assume deterministic homogeneous contact between individuals, this convenient but simplifying setting is called the mean-field (MF) assumption. Although powerful at a national scale in the early steps of an epidemic, these models show their limits (Keeling, 1999) when attempting to capture the dynamics of a heterogeneously circulating respiratory virus (Gaudart et al., 2021, **Thomine et al., 2021**), locally limited and sporadically driven by random super-spreading events (Lakdawala and Menachery, 2021). In this task, we will first use agent-based modelling and social network data analysis to explore the relevant determinants of respiratory virus heterogeneous spread and relate the observed patterns to key spatial statistics such as spatial covariance of building and human densities. This quantitative characterisation will allow us to then construct generic corrections to the MF models capable of capturing the influence of spatial structure and stochasticity (due to inter-individual transmission heterogeneity and super-spreading events), in line with previous attempts (Keeling and Rohani, 2002; Roy and Pascual, 2006). These corrections would consist in functional adjustments based on higher order spatial moments characterising the heterogeneity, following the original theoretical approach introduced in Ovaskainen and Cornell (2006); Ovaskainen et al. (2014). This approach relies on approximating the spatialisation by a scaling perturbation of the MF. North and Godfray (2017) recently applied this theory to disease spread and derived a spatial-second-order approximation of the basic reproduction number  $\mathcal{R}_0$ , namely

$$\mathcal{R}_0 = \frac{\beta \bar{S}}{\gamma} \left( 1 + \epsilon \int_{\mathbb{R}^2} \mathcal{F}(\mathcal{T}) \mathcal{F}(\text{Cov}^*[S, S]/\bar{S}^2) d\omega \right) + \mathcal{O}(\epsilon^2),$$

where the spatial heterogeneity is taken into account by the product of the Fourier transforms of the transmission kernel  $\mathcal{T}$  and the spatial covariance of locations of susceptible individuals  $S^*$ ,  $\beta$  and  $\gamma$  respectively denoting the transmission and the recovery rate. Importantly, this low-dimensionality spatial correction will allow to focus the model parametrisation on biological processes rather than mathematical sophistication, which does not ensure robustness nor public health relevance (Siegenfeld et al., 2020).

---

\*. Excluding self-transmission

### 3.2.4 Task #3 : provide key estimates and projections for public health

NPIs are essential to prevent RVI (Haug et al., 2020; Peng et al., 2022; Lind et al., 2023), especially given the waning immunity (Goldberg et al., 2022). Our collective aim is therefore to find the least restrictive and costly combination of measures sufficient to keep the circulation under control. If field epidemiology was much solicited at the beginning of the epidemic, e.g. to identify infector-infectee pairs (He et al., 2020) and super-spreading events (Lakdawala and Menachery, 2021), we lack optimal tools to analyse and anticipate RVI spread in the current context, with more contagious variants (Volz, 2023), hybrid immunity (Goldberg et al., 2022), and heterogeneous compliance to barrier gestures (Delussu et al., 2022). To take one key example, estimating the current generation time -- an essential metric in quantitative epidemiology (Britton and Scalia Tomba, 2019) -- will become increasingly difficult as epidemio-surveillance data becomes increasingly biased because of the aforementioned processes (Geismar et al., 2023). In this task, we will build an epidemiological modelling framework to infer and project both local and national dynamics of the RVI parametrised with both contact and surveillance data. This framework will then allow to explore NPI scenarios stratified by age, infectious status, or immunological status. The goal will be to identify strategies that limit socio-economic, cultural, and educational costs, while capping the morbidity, estimated in Disability Adjusted Life Year (DALY) loss, induced by RVIs.

### 3.2.5 Data and methods

SARS-CoV-2 sequences are regularly provided through the GISAID platform thanks to the EMERGEN consortium (coordinated by both the ANRS|MIE and Santé publique France), which is expected to extend to other viruses. The Cerba network of medical analysis laboratories (which has handled ca. 1/3 of the volume of SARS-CoV-2 screening in France) will provide regular RT-qPCR data screening for VOC-related mutations allowing real-time monitoring of the variant-stratified incidence. Over the course of the project, approximately thousands of respiratory samples (nasopharyngeal swabs, sputum) will be collected by the CERBA laboratory network and tested for SARS-CoV-2, RSV, influenza A and B in a standardised manner (multiplex PCR). Testing data will be enriched by patient covariates allowing to correct for screening biases, with general linear mixed models and survival analysis.

The distribution cycle threshold (Ct) data provided by qPCR testing moreover constitutes an early signal of the rebound of the epidemic which can be exploited using temporal autoregression to improve short-term predictions (Alizon et al., 2022). Current surveillance data will be enriched by the Infectious Diseases Division of Santé publique France past epidemio-surveillance of RVIs to decipher the influence of SARS-CoV-2 co-circulation and related measures from inter-annual variations. In order to highlight positive or negative intra-host interactions data from multiple infection data, I will apply a recent mechanistic model-based inference method developed by the team (Alizon et al., 2019), along with non-supervised ma-

chine learning algorithms (e.g. linear tools such as Lasso regression or non-linear tools such as random forests).

To analyse the spatial spread of the RVIs, we will use the Epidemap (Thomine et al., 2021) agent-based modelling (ABM) software. Epidemap allows realistic simulation of the spread of one or more pathogens within a heterogeneous population, both spatially (habitat structure imported from OpenStreetMap) and stochastically (the millions of inhabitants of a city or country are explicitly represented and move from one building to another every day). The functional determinants of spatial propagation will also be made possible by studying social network data (the *Data for Good* partnership is already established between *Facebook* and myself via the University of Montpellier).

Case-contact and cluster data will be provided by the Agences Régionales de Santé (ARS) Occitanie region, via its joint Surveillance Outbreak Response Management & Analysis System (SORMAS) platform (Fähnrich et al., 2015). After a text mining preliminary step, multivariate statistical analyses and machine learning techniques (artificial neural networks, random forests, classification and regression tree) will be performed to identify prevention gaps and effective measures.

These updated values will help parametrise the core framework of the project combining two previously developed models. The first one was designed to monitor the dynamics of SARS-CoV-2 in France (Sofonea et al., 2021c) and is capable of capturing hospital activity more than a month in advance thanks to its non-Markovian approach (Sofonea and Alizon, 2021) (see above and fig. 3.6). The second is a nested epidemiological coinfection model with explicit within and between-host dynamics that can feature any number of co-circulating strains, variants and pathogen species (Sofonea et al., 2015), the generic structure of which is shown in fig. 3.12. The synthetic framework will take into account the MF corrections of Task 2 and the surveillance of SARS-CoV-2 variants and other respiratory viruses from Task 1. Finally, numerical optimisation will be applied to this non-Markovian multiple infection model to seek for the lightest public health policies which limit the winter epidemic of RVIs.



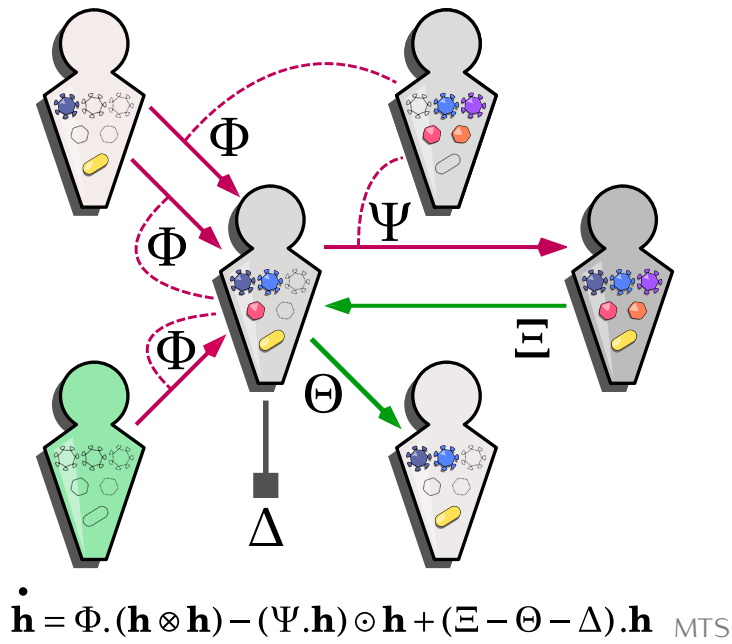


FIGURE 3.12 – **Structure of the set-based generic multiple infection model.**

The epidemiological flow chart focuses on a central host class represented as an arbitrary set of four out of six distinct pathogen strains, variant and species. Only a subset of host classes and possible transitions are depicted so that each type of event is shown at least once. Each flow is associated with a greek letter which denotes the corresponding matrix in the master multivariate ordinary differential equation (ODE) of the model, shown at the bottom. Infection, recovery and removal flows are respectively depicted by pink, green and grey arrows. Adapted from Sofonea et al. (2015).

## 3.3 Perspectives à long terme

### Portraying Pathogen X : from fundamental microbiology to meta-epidemiology (P2X) \*

#### 3.3.1 Context

In three years, the SARS-CoV-2 pandemic will have taken nearly seven million lives according to official counts (WHO, 2023a) (but potentially three times as many (Wang et al., 2022)), and caused ten times as many post-acute sequelae (Davis et al., 2023). Beyond its health burden, the crisis has caused lasting socio-economic and cultural damage, estimated at several tens of billions of euros for a country like the United Kingdom, having amplified social inequalities (Perry et al., 2021; Oberndorfer et al., 2022), but also degraded citizens' confidence in science (Ipsos, 2022).

The emergence of SARS-CoV-2 is not an isolated event. It is part of the list of emerging and re-emerging viral infectious diseases observed over the last two decades : dengue, SARS, Chikungunya, Rift Valley fever, pandemic influenza, Ebola, Zika, yellow fever, Nipah (Meganck and Baric, 2021), and more recently Mpox (ECDC, 2022). In this context, the scientific and sanitary preparedness of states for a new health crisis of infectious origin appears crucial and urgent. In its list of the main priority pathogens for research and development relating to health crises, the World Health Organisation introduced, in February 2018, the mention of a *Disease X*, specifying that it "represents the knowledge that a serious international epidemic could be caused by a pathogen currently unknown to cause human disease" (WHO, 2018). Following the COVID-19 pandemic – which would have corresponded to Disease X up to 2020 – the WHO recently renewed this mention (in November 2022), thus encouraging the international community to extend research on infectious diseases and their control beyond the scope of the pathogens documented to date (Van Kerkhove et al., 2021). Unfortunately, in spite of its importance, the definition of Disease X is currently too imprecise to be actionable from a research or public health perspective. The literature on the subject is essentially composed by a few opinions (but see Simpson et al. (2020) on the topic of biological preparedness). The ultimate ambition of this project is to provide the yet lacking meaning and usefulness to the notion of Disease X.

---

\*. Ce projet, co-porté avec Pr Sonia Burrel (U. Bordeaux & CHU Bordeaux, ex-Sorbonne U. & AP-HP et présidente de la Société Française de Microbiologie) correspond au *workpackage* central d'un projet plus important (*Preparing research, care & Public Health policies for Pathogen X, PrePHiX*), que j'ai déposé au printemps 2023 à l'appel d'offres [PEPR Maladies infectieuses émergentes](#). Réunissant neuf équipes françaises de recherche en virologie, infectiologie, phylogénomique, santé publique et sciences sociales autour de la préparation pandémique, le projet PrePHiX a déjà reçu le soutien de la [Société Française de Microbiologie \(SFM\)](#), de la [Société de Pathologie Infectieuse de Langue Française \(SPILF\)](#), de la [Société Française d'Anesthésie-Réanimation \(SFAR\)](#), de l'[Infectious Diseases Models for Innovative Therapies](#) (infrastructure du Commissariat à l'énergie atomique et aux énergies alternatives), du [Vaccine Research Institute \(ANRS | MIE\)](#) et de [Santé publique France](#).

### 3.3.2 Goals

With a view to pandemic preparedness, the profile of *Pathogen X*, the aetiological agent of Disease X, has yet to be drawn. To this end, we will first implement a fundamental and etiology-driven approach, which consists in reasoning from the already documented microbiological world and locating its most plausible origins and features in the pathosphere (Randles, 1999). We will also develop a more applied and public health-driven approach, which consists in identifying the limits and weaknesses of the healthcare system and the latency of the scientific, surveillance, and prevention resources that can be mobilised to deal with an emerging infectious disease. These elements will allow us to define the features a new pathogen may have that would put our healthcare system under pressure or even in danger. The aim of this project will be to implement these approaches using the most advanced qualitative and quantitative tools and methods. Resolutely at the interface between fundamental microbiology, infectiology, public health, and modelling, this work will develop an original approach that will address both fundamental and applied questions of major scientific and health interest. The project outline and main projected inputs, interactions, and deliverables, are shown in fig. 3.13, and developed below.

### 3.3.3 Task #1. Identifying Pathogen X relevant features via crowdsourced expert judgment.

For more than forty years, the most worrying pandemics and outbreaks – from AIDS to Ebola virus disease (EVD) to COVID-19 – were all caused by RNA viruses (Meganck and Baric, 2021), with the only notable exception to the Mpox outbreak of the spring and summer of 2022, which was caused by a DNA virus. It is also now well established that more than two-thirds of emerging infectious diseases are of zoonotic origin, whose host species switching is facilitated by anthropogenic activities affecting wild ecosystems (Destoumieux-Garzón et al., 2018). Rapid pandemic spread is then made possible by the combination of human-to-human transmission of the agent (aerosols, droplets, sexual and so-called direct contact), and the human population density of urban centres and interconnection of urban centres (both from the point of view of topology and flows (Zhang et al., 2020)).

All of the above are the features plausibly attributed to Pathogen X in the limited literature on the topic (Simpson et al., 2020). Unfortunately, these criteria are not specific enough. Indeed, three new RNA viruses capable of infecting humans are described each year, on average (Woolhouse and Brierley, 2018). To avoid triggering false alerts that might do more harm than good, new criteria must be established (Woolhouse et al., 2016).

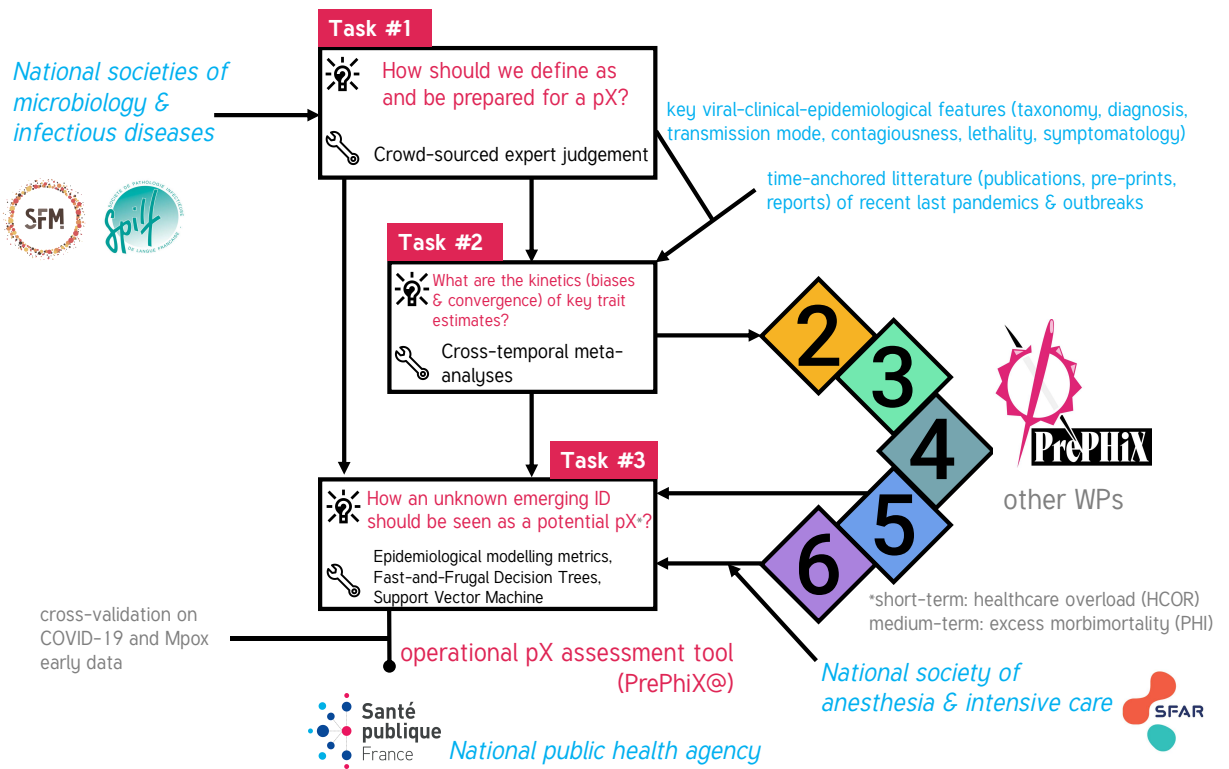


FIGURE 3.13 – Schematic organisation of the P2X project.

Each of the three connected tasks addresses distinct questions with proper methods and receives different kinds of external inputs, either from learned societies, systematic review, or parallel work from the other workpackages of the PrePHiX master project. Besides a PhD student to be recruited, the personnel involved in the research include Mircea T. Sofonea (epidemiology, evolutionary biology, biostatistics – U. Montpellier), Pr Sonia Burrel (medical virology – U. Bordeaux & Bordeaux Univ. Hosp.), Pr Édouard Tuaille & Pr Vincent Foulgoune (medical virology – U. Montpellier & Montpellier Univ. Hosp.), Pr Yannick Simonin (fundamental virology – U. Montpellier) and Dr Sylvie Abel & Pr André Cabié (infectiology – U. Antilles & Fort-de-France Univ. Hosp). WP : workpackage, pX : Pathogen X.

One solution to refine the portrait of Pathogen X is to use epidemiology, intuitively looking at severity/fatality and contagiousness (calculated as the number of basic reproductions,  $\mathcal{R}_0$ ). figure 3.14 maps some of the priority pathogens (WHO, 2018) in this reference frame, using current estimation ranges. These cover a wide area, from low ( $\mathcal{R}_0 < 1$ ) to high ( $\mathcal{R}_0 \geq 3$ ) contagiousness, and several orders of magnitude of lethality (from 0.1% to 90%). Therefore, additional features have to be considered. For example, Fraser et al., 2004 proposed to also include the proportion  $\theta$  of transmission events that occur in the absence of symptoms (asymptomatic infectors or during the incubation period), as a proxy for epidemic uncontrollability. In the case of SARS-CoV-2, the  $\theta$  early estimate was 44% (95% confidence interval (CI), 30-57%) (He et al., 2020).

To identify a maximum of relevant features - and their interrelationships - for the definition of Pathogen/Disease X, we will call upon the collective expertise of professions with historical experience, reinforced by the COVID-19 pandemic, namely microbiologists and infectious diseases specialists. This will be done through the French society of microbiology (SFM, chaired by Pr Sonia Burrel) and the French-speaking society of infectious pathology (SPILF). Their large network will be asked to circulate a questionnaire (via an anonymous online survey platform) on the definition of Pathogen X and pandemic preparedness. The questionnaire will be developed in collaboration with representatives of the two learned societies prior to submission and virologist and infectious disease clinicians of the PCCEI research unit (Edouard Tuillon, Vincent Foulogne, Yannick Simonin in Montpellier, André Cabié and Sylvie Abel in Fort-de-France (French West Indies)). Most of the questions will be multiple choice with the possibility to rank (possibly with ties) and to add suggestions.

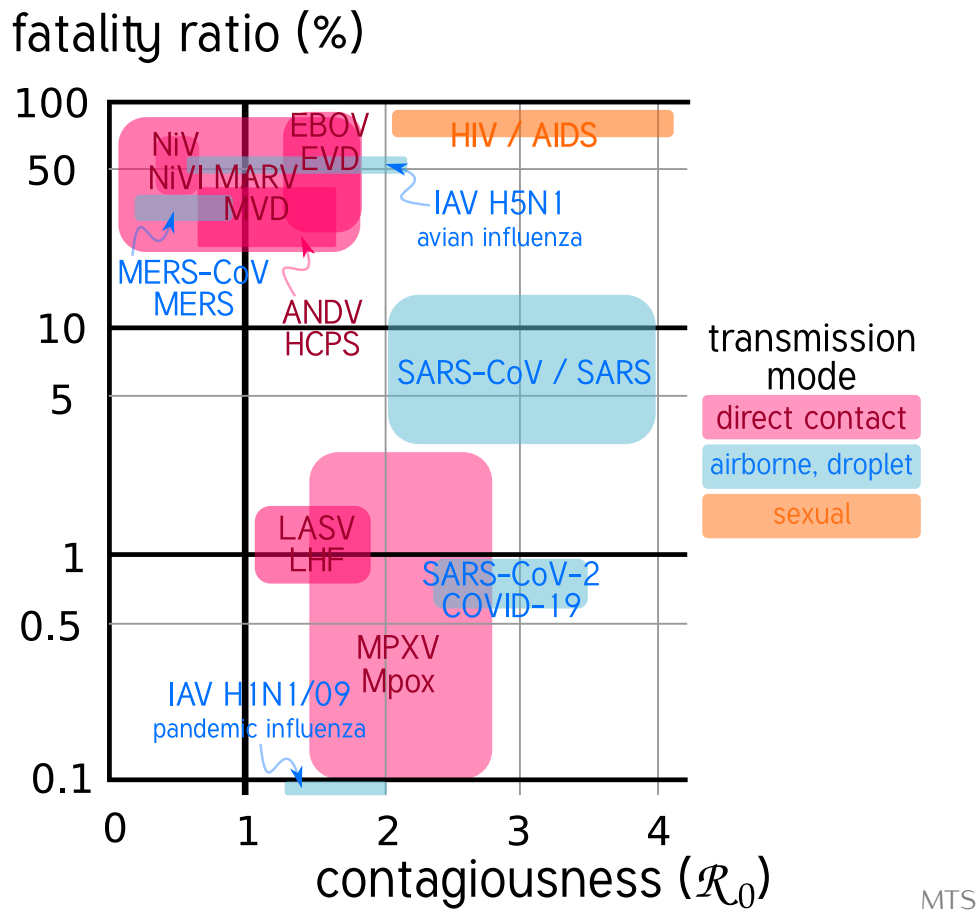


FIGURE 3.14 – Severity & contagiousness mapping of the major past or present emerging pathogens & infectious diseases.

The severity is provided, in logarithmic scale, as estimated ranges of fatality ratios (among infections or cases). The contagiousness is measured by the range of initial estimates of the basic reproduction number ( $\mathcal{R}_0$ ). The colour depicts the main transmission route. The values were searched in World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), and European Centre for Disease Prevention and Control (ECDC) factsheets (e.g. (CDC, 2019; ECDC, 2022; WHO, 2023b) or from the epidemiological literature (e.g. Sofonea et al., 2021c). AIDS : Acquired immunodeficiency syndrome. ANDV : Andes orthohantavirus. COVID-19 : Coronavirus disease 2019. EBOV : Ebola virus. EVD : Ebola virus disease. HCPS : Hantavirus cardiopulmonary syndrome. HIV : human immunodeficiency virus. IAV : Influenza A virus. LASV : Lassa mammarenavirus. LHF : Lassa haemorrhagic fever. MARV : Marburg virus. MERS : Middle East respiratory syndrome. MERS-CoV : Middle East respiratory syndrome-related coronavirus. MPXV : Mpox virus. MVD : Marburg virus disease. NiV : Nipah virus. NiVI : Nipah virus infection. SARS : Severe acute respiratory syndrome. SARS-CoV(-2) : Severe acute respiratory syndrome coronavirus (2).

Personal opinion will be sought on e.g. which microbiological groups (at family and genus/species level) Pathogen X might belong to? Which qualitative (which tissue tropism, which symptomatology) and quantitative (mean incubation period, reproduction number) features should be included in the definition of Pathogen X? What should the corresponding thresholds be? What microbiological and infectious features of Pathogen X could put our research, surveillance and care system under strain or even degrade it and against which it requires enhanced preparedness? For example, the experience of the triple epidemic of respiratory viruses (COVID-19, influenza, bronchiolitis) in the 2022-2023 autumn-winter season has highlighted the overload risk of paediatric critical care (Roucaute, 2023). A pathogen that is too frequently severe in children would thus fall within the scope of Pathogen X. Similarly, a disease with severe renal damage would rapidly pose a risk of hospital saturation due to the need for dialysis and longer hospital stays.

This approach to soliciting collective expertise is called crowdsourced human judgment and has been used profitably (via platforms such as *Metaculus*) in situations of uncertainty such as the COVID-19 pandemic projections. At the beginning of the Mpox outbreak, they also proved to be a source of signal that deserves to be further exploited in public health (McAndrew et al., 2022).

The descriptive and inferential analysis of the survey responses will provide an overview, stratified by profession, of expert opinion on the concept of Disease and Pathogen X. It will inform and focus the work of the next two tasks and provide *a priori* distributions for the work of the other work packages of the PrePHiX project.

### 3.3.4 Task #2. Key trait joint distribution and estimation kinetics

Some of the key quantitative features (hereafter traits) we expect to find from Task #1 are the basic reproduction number, the average lethality, the generation time or serial interval, the incubation period, or the fraction of asymptomatic infections. To provide Pathogen X with a multivariate (joint) probability distribution of its potential trait values, the corresponding estimates of the proxy pathogens it would most resemble must be updated and robust. However, the significant differences in estimates of contagiousness and apparent lethality of priority pathogens (see fig. 3.14) from one source to another show a lack of numerical consensus. As the emblematic case of measles  $\mathcal{R}_0$  illustrates, the variance of estimates can be significant depending on the spatial and temporal origin of the data and the methods used (Guerra et al., 2017). Some historical values have also become obsolete (Delamater et al., 2019), if only because, in the case of fatality ratios, demographics and management (even if only symptomatic) have changed. To our knowledge, the few references gathering the key trait estimates of priority pathogens are incomplete and potentially outdated (Wolfe et al., 2007). A first output of this task will be to collate meta-analyses, if available, of the values of these traits, for the pathogens identified in the previous task as proxies of Pathogen X.

While the traits of a new emerging pathogen can be used to identify it as a Pathogen X (ob-

jective of Task #3), early estimates may suffer from bias and imprecision. For example, the estimates of the incubation period of COVID-19 by date of online availability (peer-reviewed or preprints) in fig. 3.15 show, as a preliminary result, that the average duration has increased by almost 50% among studies in 5 months. The confidence intervals also narrowed over time. The underestimation bias of inter-event delays in a growing epidemic is well known in the literature (Britton and Scalia Tomba, 2019). However, its systematic correction is not always applied or applicable. In the case of incubation period, it leads to an overestimation of the controllability of the epidemic (by underestimating the fraction of a- and pre-symptomatic transmission) and a reduction in the effectiveness of contact monitoring and isolation, by suboptimally designed time windows.

Analogously with the COVID-19 pandemic, where it took about 200 days for the first wave to pass in most countries of the world and for the estimates to be consolidated, we will conduct a systematic literature review of estimates (preprints, publications, and technical reports) of key parameters of Pathogen X proxies, over the first 200 days of their outbreak. A cross-temporal meta-analysis (CTMA, (Karadag, 2020)) of the time-anchored estimates compared to the consolidated estimates will reveal, for each parameter, the direction and magnitude of the bias, if any, as well as the kinetics of its reduction, and that of the uncertainty. On the basis of early field data sets for COVID-19 and Mpox, heuristic bias correction and refinement methods combining machine learning (ETS, ARIMA, AR-Lasso (Hyndman and Athanasopoulos, 2018)), data augmentation (Cooper et al., 2008) and data imputation (Sportisse et al., 2020) will be developed and tested by cross-validation. All these analyses and tools will be developed in R using the `meta`, `fable`, `augmenterR` and `missMDA` (Josse and Husson, 2016) packages.



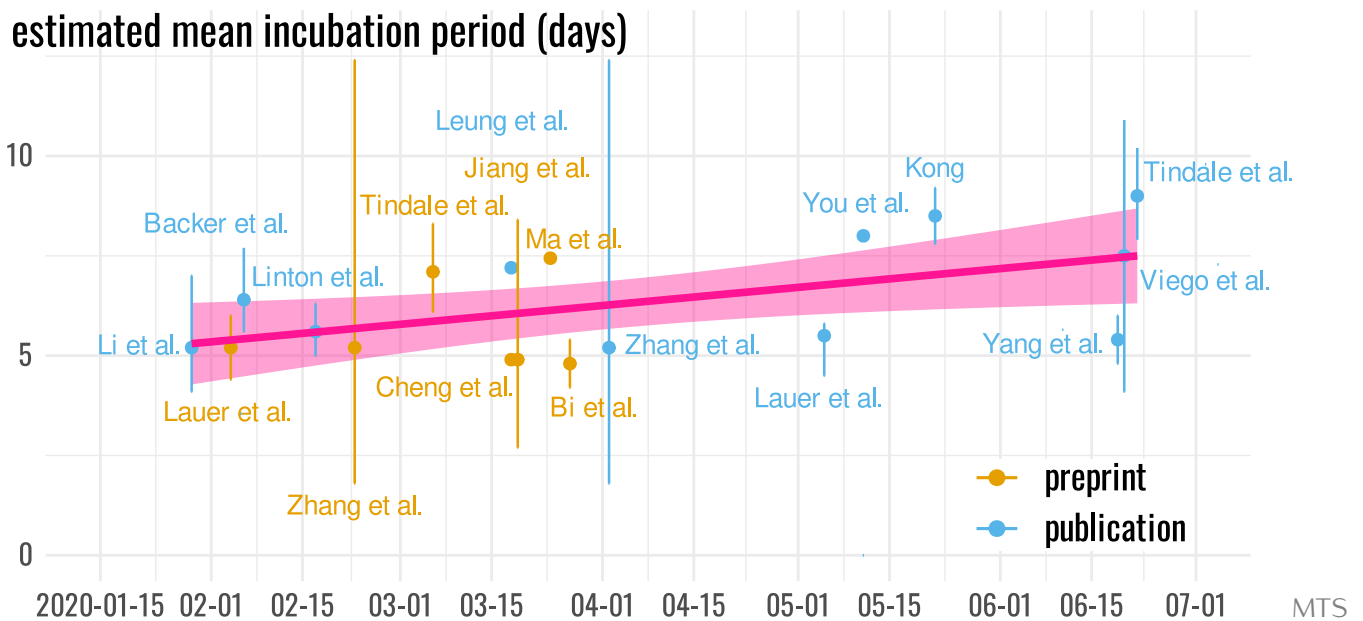


FIGURE 3.15 – **Incubation period estimates of SARS-CoV-2 infections in the first six months of the COVID-19 pandemic.**

The estimated means of the COVID-19 incubation period and their 95 % confidence interval (CI) are plotted along the date of their online availability. First author's last names are indicated above the estimates. Preprints are shown in orange and peer-reviewed articles in blue. The pink line and ribbon indicate the linear regression of the estimated means as a function of online availability date. The selection of the studies follows that of the incubation period meta-analyses of McAloon et al. (2020) and Alene et al. (2021).

### 3.3.5 Task #3. Design complementary Pathogen X assessment metrics and tools using modelling and machine learning

The final task of this project will be to operationalise the previously enriched definition of Pathogen X. As discussed in Task #1, contagiousness and severity will not be sufficient to classify a newly identified pathogen as Pathogen X, in the sense of the need for health countermeasures. By knowing additional key parameters, or by inferring them by augmentation and imputation methods from the data discussed in Task #2, it will be possible to provide enough signal to discriminate the pathogen as a sanitary concern. We will thus develop two new metrics that capture the effect of Pathogen X on the medium term and on the short term respectively.

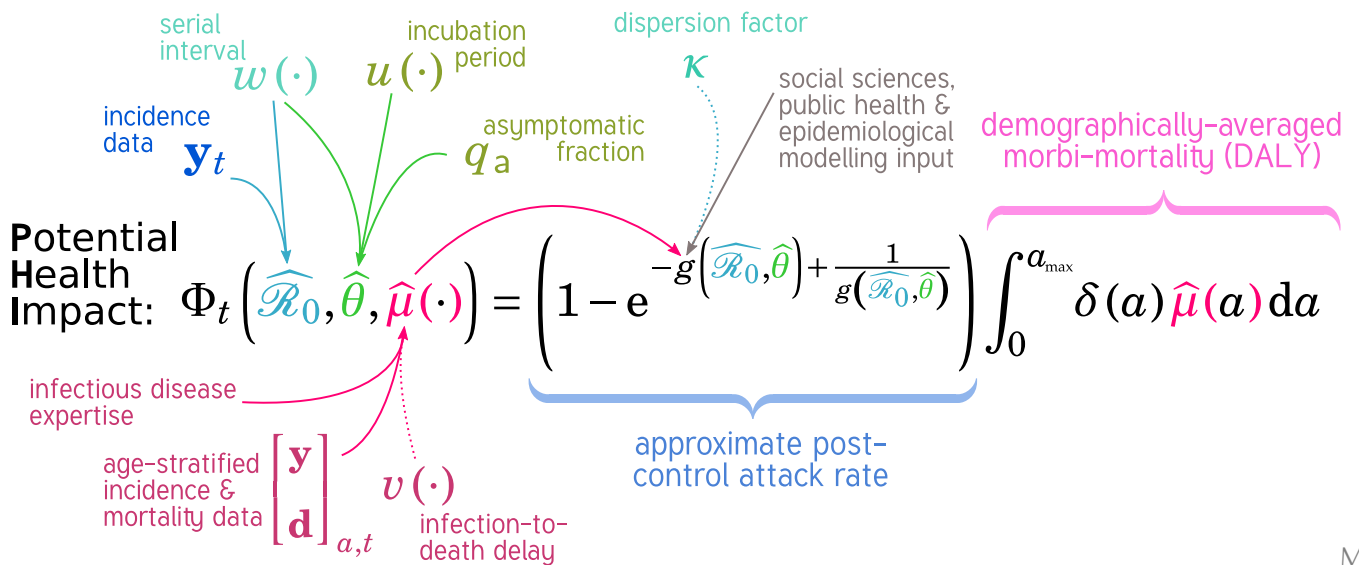
First, the potential health impact, PHI ( $\Phi$ ), aims to quantify, early in the pandemic, a health burden in disability-adjusted life years (DALYs), which will be weighted by a projected attack rate. A first version of its formula, as a working basis, is given in fig. 3.16. This metric will be co-constructed with *Santé publique France* (SpF), the national public health agency. Several control scenarios will be explored, building on the results of the behaviour/social sciences WP of the PrePHiX project (through the modulation of the  $g$  function, see caption in fig. 3.16). Once consolidated, the  $\Phi$  metric will be applied to the COVID-19 and Mpox data over the first 200 days and its successive values compared to the a posteriori observations. Similar to the

work carried out in Task #2, the possible biases and convergence kinetics observed on these two recent epidemics will allow phenomenological correction of this metric, while providing benchmarks against which to compare the values of future emerging pathogens.

Second, the healthcare overload risk, HCOR, will be derived from the results produced by WP6 of the PrePHiX project (Decision-making & hospital preparedness), itself building on the results generated by the other WPs (see fig. 3.13). In addition, it will integrate the updated critical care capacity data provided by the French society of anaesthesia and intensive care (SFAR). This metric will be standardised in the first wave of COVID-19 in France in order to provide a transparent and evocative reference for all the country's health actors and decision-makers.

In the very first weeks of the outbreak (whether the origin is located in France or abroad), the metrics PHI and HCOR might however not be computable. This can be due to lack of data, e.g. the absence of reliable infector-infectee pairs prevent the estimation of the basic reproduction number, or unobserved cases among some age classes hinders profiling the hospitalisation and lethality probability across ages. Following the guidelines made by Zelner et al. (2021), the propagation of the uncertainty through the successive steps of the metric computation, including the lack of consensus among experts on the countermeasure parametrisation, could also generate confidence intervals too wide to be relevant for decision-making. To address this issue, robust automated classification methods need to be explored as alternatives. Two machine learning approaches will be used : Fast-and-Frugal decision trees, in case of very limited data, and Support Vector Machines, for better documented emerging diseases. Their principle and application to pandemic preparedness are detailed in a dedicated box hereafter.

The calculation of PHI, HCOR and the two classification algorithms will be implemented in R using original scripts (for the metrics) and dedicated packages (FFTrees (Phillips et al., 2017) for the FFTs and e1071 (Dimitriadou et al., 2006) for the SVMs). These four methods of evaluation of Pathogen X will be brought together in a single operational tool, the PrePHiX@ (for Preparing research, care & Public Health policies for Pathogen X – Assessment Tool) graphically interfaced in R Shiny, to allow for the implementation by SpF in the first place but also the French regional health agencies (ARS), the high health authority (HAS), the high public health council (HCSP) and the ministry of health services (DGS). PrePHiX@ will be maintained over the long term by the collaborators of the PrePHiX project and jointly with *Santé publique France*. The principle of the tool will be disseminated to promote its transportability, in particular to low- and middle-income countries (LMICs) and countries with hot spots of infectious emergencies (Destoumieux-Garzón et al., 2018).



MTS

FIGURE 3.16 – Calculation of the Potential Health Impact metric.

The Potential Health Impact (PHI,  $\Phi$ ) is an aggregated metric aimed to rapidly assess the health burden of a potentially-mitigated emerging pathogen, provided current estimates (denoted by hats) of key traits at time  $t$  since the notification of the first world case (i.e. the outbreak age). (NB : For the sake of simplicity, the dependence of the estimates to the outbreak age is not denoted). In a first, naive, formulation,  $\Phi$  would only take three arguments – the reproduction number  $\mathcal{R}_0$ , the a- and pre-symptomatic transmission fraction  $\theta$ , and the morbi-mortality distribution across ages  $\mu(\cdot)$  – and be the product between an attack rate (in  $(0, 1)$ ) and an average morbi-mortality quantity upon infection (scaled in disability-adjusted life years, DALY). The data involved in the computation of each of the three arguments (usually originating from field epidemiology and infectious-disease expertise from microbiologists and clinicians) are detailed at the start of the incoming arrows (the dotted arrows are not required for crude estimations).

The attack-rate factor (blue under-brace) assumes no contact heterogeneity across ages (uniform contact matrix) and uses a heuristic approximate solution of the Markovian mean-field demographically homogeneous Kermack & McKendrick’s final epidemic size equation (Miller, 2012). Importantly, it involves a controllability function  $g(\cdot)$ , which lowers the unmitigated value of  $\mathcal{R}_0$  if there is leverage for control. E.g., in a simplistic scenario where the only countermeasure implemented is a perfect prevention of transmission from symptomatic individuals, then  $g(\mathcal{R}_0, \theta) = \theta \mathcal{R}_0$ . More realistic formulations of  $g(\cdot)$  will be considered using modelling (provided by the other WPs of the Pre-PHiX project) to synthesise the infectious disease, social science and public health inputs (e.g. what would be the compliance to mask wearing depending on the symptomatology of Disease X?), including the dispersion factor  $\kappa$  quantifying the contribution of super-spreading.

The morbi-mortality factor (pink over-brace) is computed by integrating the age-stratified DALY upon infection  $\mu(\cdot)$  between 0 and maximal age  $a$ , weighted by the age pyramid distribution  $\delta(\cdot)$  currently provided a projected for next four decades by (Algava and Blanpain, 2021).

The quantiles (median and confidence intervals) of  $\Phi$  will be generated by the Monte-Carlo method, using independently randomly sampled values from the inferred distribution of each key parameter.

**Fast-and-Frugal decision trees** (FFTs) are explicit decision-making tools tailored for uncertain and resource-constrained conditions. They are heuristics-based models that help make accurate predictions or classifications by using limited signal in a dichotomous step-wise manner, facilitating timely decisions (Hafenbrädl et al., 2016). FFTs are generated through a systematic process that involves the identification and selection of relevant information (from both empirical data and expert knowledge), followed by the determination of the order in which these cues are considered, and finally, the establishment of decision thresholds for each cue. The resulting decision tree is composed of an ordered sequence of binary decision nodes, each representing a cue, and leading to either a decision outcome or another cue to consider. In the context of pandemic preparedness, FFTs can be particularly helpful in the early stages of a pandemic since they allow for rapid decision-making in situations where, in early stages, data is scarce and time is critical while their simplicity and transparency enable clear communication of decision-making processes to stakeholders, fostering trust and cooperation (Hafenbrädl et al., 2016).

**Support Vector Machines** (SVMs) are a class of supervised machine learning algorithms that can be used for classification or regression (Noble, 2006). They are particularly effective in high-dimensional spaces and work especially well when the number of traits is larger than the number of samples. Pandemic preparedness exactly falls into the latter situation since the number of recent pandemics and concerning outbreaks is small compared to the number of biological, clinical and epidemiological features (collected from Tasks #1 and #2, potentially stratified by country and wave) that can be assigned to each of them. SVMs function by finding the optimal hyperplane that maximises the margin between different classes, thus ensuring the best separation of data points. SVMs are generated by training a model using labelled data, where each instance is associated with a class label. The algorithm searches for the optimal hyperplane by minimising an objective function that considers the margin width between different classes and the misclassification errors. This is achieved through the use of kernel functions, which allow the algorithm to transform the input space into a higher-dimensional space where the data points can be more easily separated, as illustrated by fig. 3.17. Once the model is trained on past outbreak data (from WHO, CDC, ECDC or SpF databases) and key trait estimates (Task #2), the SVM can be used to classify the threat posed by any newly emerged pathogen through the position of its corresponding point with respect to the learned decision boundary (Pourhomayoun and Shakibi, 2021).

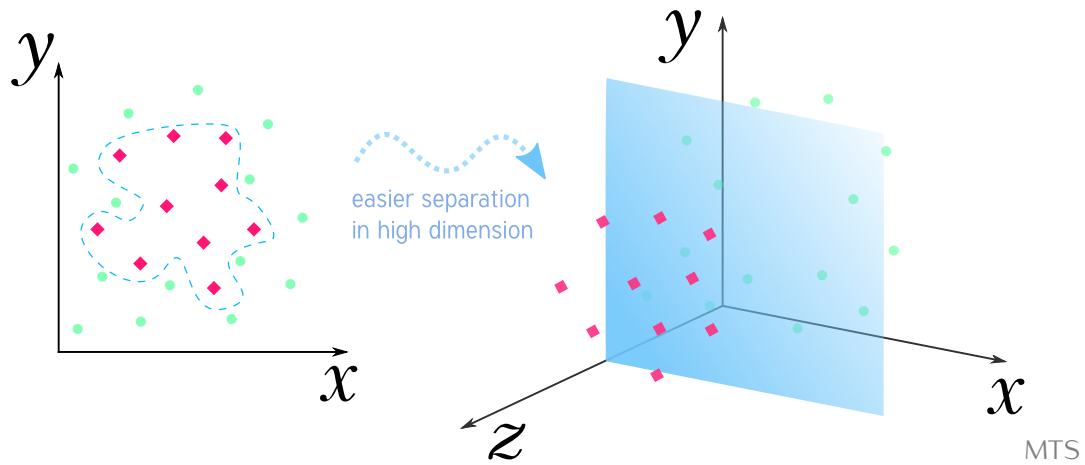


FIGURE 3.17 – **Principle of Support Vector Machines.**

Public-health-system threatening pathogens are depicted as pink diamonds. Less concerning pathogens are represented by green dots. In a two-dimensional frame (left) such as the severity-contagiousness frame (fig. 3.14), the Pathogen X delineation would be a highly-non linear separatrix (blue dashed curve), therefore difficult to infer without risking misclassification. Support Vector Machine are supervised learning models that can circumvent this issue by mapping data into spaces of higher dimensions (a technique known as the kernel trick). Adding new, possibly aggregated, traits through optimised transformation, the points become readily separable by a hyperplane, the equation of which provides a more operational definition of the Pathogen X.

### 3.3.6 Impact

From scientific and public health perspectives, this project will allow to integrate multidisciplinary collective expertise, the fragility of the healthcare system, and the latency of research, surveillance, and prevention, by means of cutting-edge qualitative and quantitative approaches, so as to provide a tool for the early assessment of the health threat of future emerging pathogens, brought together under the label of Pathogen X thus identified and actionable. In terms of human resources, this project will allow for the high-level training of a PhD student who will reinforce the demanding and highly interdisciplinary scientific skills in pandemic preparedness.

4

# Annexes

## 4.1 Publications sélectionnées

référence	page
<p><b>[10]</b> Sofonea MT, <a href="#">Reyné B</a>, Elie B, Djidjou-Demasse R, Selinger C, Michalakis Y, Alizon S. (2021) Memory is key in capturing COVID-19 epidemiological dynamics* . <i>Epidemics</i>. 35, 100459</p>	<p>84 (&amp; annexes)</p>
<p><b>[17]</b> Alizon S, Haim-Boukobza S, Foulongne V, Verdurme L, Trombert-Paolantoni S, Lecorche E, Roquebert B, <b>Sofonea MT</b>. (2021) Rapid spread of the SARS-CoV-2 Delta variant in French regions in June 2021 † . <i>Eurosurveillance</i>. 26(28) :pii=2100573.</p>	<p>92 (&amp; annexes)</p>
<p><b>[18]</b> <b>Sofonea MT</b> &amp; Alizon S. (2021) Anticipating COVID-19 intensive care unit capacity strain : A look back at epidemiological projections in France. <i>Anaesthesia Critical Care &amp; Pain Medicine</i>. 40(4), 100943.</p>	<p>98</p>
<p><b>[21]</b> Thomine O, Alizon S, Barthelemy M, <a href="#">Boennec C</a>, <b>Sofonea MT</b>. (2021) Emerging dynamics from high-resolution spatial numerical epidemics. <i>eLife</i>. 10 :e71417.</p>	<p>102 (&amp; annexes)</p>
<p><b>[30]</b> <b>Sofonea MT</b>, Roquebert B, Foulongne V, Morquin D, Verdurme L, Trombert-Paolantoni S, Roussel M, Bonetti JC, Zerah J, Haim-Boukobza S, Alizon S. (2022) Analyzing and modeling the spread of SARS-CoV-2 Omicron lineages BA.1 and BA.2 in France (September 1, 2021, to February 28, 2022). <i>Emerging Infectious Diseases</i>. 28(7) :1355-1365.</p>	<p>116 (&amp; annexes)</p>

\*. Cette publication correspond à la sous-partie majoritairement méthodologique du *preprint* mis en en ligne en mai 2020 (**Sofonea et al., 2020b**, dont les autres résultats n'ont jamais été publiés). Elle introduit les fondations théoriques et computationnelles des logiciels COVIDSIM-FR (**Sofonea et al., 2020a**) et COVIDici (**Boennec et al., 2021b**), qui furent utilisé dans la projection nationale, régionale et départementale de l'activité COVID-19 hospitalière. Ce cadre de travail a été adapté par la suite aux vagues suivantes, et transposée à l'épidémie sur l'île d'Irlande (**Kamiya et al., 2023**, **Fitz-Simon et al., 2023**).

†. Cette publication constitue la première estimation de la surcontagiosité du variant Delta en France métropolitaine. Sa version préliminaire a servi de support précoce à la Direction Générale de la Santé (DGS, Ministère de la Santé).

# Bibliographie

- Adam D. Special report : The simulations driving the world's response to COVID-19. *Nature* 2020 Apr;580(7803) :316–318. **38, 60**
- Akay R, caretForecast - Conformal Time Series Forecasting Using State of Art Machine Learning Algorithms. CRAN; 2022. <https://github.com/Akai01/caretForecast>. **55**
- Alene M, Yismaw L, Assemie MA, Ketema DB, Gietaneh W, Birhan TY. Serial interval and incubation period of COVID-19 : a systematic review and meta-analysis. *BMC Infectious Diseases* 2021 Mar ;21(1) :257. **73**
- Algava E, Blanpain N. 68,1 millions d'habitants en 2070 : une population un peu plus nombreuse qu'en 2021, mais plus âgée. *Insee Première* 2021 Nov;(1881). **75**
- Alizon S, Haim-Boukobza S, Foulongne V, Verdurme L, Trombert-Paolantoni S, Lecorche E, et al. Rapid spread of the SARS-CoV-2 Delta variant in some French regions, June 2021. *Eurosurveillance* 2021 Jul ;26(28) :2100573. **39, 46, 54, 57, 61**
- Alizon S, Murall CL, Saulnier E, Sofonea MT. Detecting within-host interactions from genotype combination prevalence data. *Epidemics* 2019 Jun ;p. 100349. **38, 63**
- Alizon S, Selinger C, Sofonea MT, Haim-Boukobza S, Giannoli JM, Ninove L, et al. Epidemiological and clinical insights from SARS-CoV-2 RT-PCR crossing threshold values, France, January to November 2020. *Eurosurveillance* 2022 Feb ;27(6) :2100406. **56, 63**
- Alizon S, Sofonea MT. SARS-CoV-2 virulence evolution : Avirulence theory, immunity and trade-offs. *Journal of Evolutionary Biology* 2021 Jul ;00 :1–11. **57, 61**
- Anderson RM, Vegvari C, Hollingsworth TD, Pi L, Maddren R, Ng CW, et al. The SARS-CoV-2 pandemic : remaining uncertainties in our understanding of the epidemiology and transmission dynamics of the virus, and challenges to be overcome. *Interface Focus* 2021 Oct ;11(6) :20210008. **60**
- Andreasen V, Lin J, Levin SA. The dynamics of cocirculating influenza strains conferring partial cross-immunity. *Journal of Mathematical Biology* 1997 ;35(7) :825–842. **61**
- Arifin SMN, Madey GR, Collins FH. Spatial Agent-Based Simulation Modeling in Public Health : Design, Implementation, and Applications for Malaria Epidemiology. John Wiley & Sons ; 2016. **39**
- Beneteau T, Elie B, Sofonea MT, Alizon S. Estimating dates of origin and end of COVID-19 epidemics. *Peer Community Journal* 2021 ;1. **41, 51**
- Beneteau T, Selinger C, Sofonea MT, Alizon S. Episome partitioning and symmetric cell divisions : Quantifying the role of random events in the persistence of HPV infections. *PLOS Computational Biology* 2021 Sep ;17(9) :e1009352. **38**



- Berrod N. Déconfinement : la piste du "stop and go" envisagée par le gouvernement. *Le Parisien* 2020 Apr; . 49
- Bhatia S, Imai N, Watson OJ, Abbood A, Abdelmalik P, Cornelissen T, et al. Lessons from COVID-19 for rescalable data collection. *The Lancet Infectious Diseases* 2023 Sep ;23(9) :e383–e388. 61
- Bjornstad ON. *Epidemics : Models and Data Using R*. Springer Nature; 2022. Google-Books-ID : doOhEAAAQ-BAJ. 39
- Boennec C, Alizon S, Sofonea MT, Anticipating the hospital burden of future COVID-19 epidemic waves. medRxiv; 2021. <https://www.medrxiv.org/content/10.1101/2021.08.19.21262280v1>, pages : 2021.08.19.21262280. 33
- Boennec C, Alizon S, Sofonea MT, COVIDici - Visualization & projection tool of the COVID-19 populational data in France at national, regional and departmental levels. France Bioinformatique; 2021. <https://cloudapps.france-bioinformatique.fr/covidici/>. 33, 53, 82
- Boennec C, Alizon S, Sofonea MT, COVimpact - Simulator of the hospital burden due to COVID-19. France Bioinformatique; 2021. <https://cloudapps.france-bioinformatique.fr/covimpact/>. 33
- Bonacina F, Boëlle PY, Colizza V, Lopez O, Thomas M, Poletto C. Global patterns and drivers of influenza decline during the COVID-19 pandemic. *International Journal of Infectious Diseases* 2023 Mar ;128 :132–139. 60
- Bouadma L, Lescure FX, Lucet JC, Yazdanpanah Y, Timsit JF. Severe SARS-CoV-2 infections : practical considerations and management strategy for intensivists. *Intensive Care Medicine* 2020 Apr ;46(4) :579–582. 44
- Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5 : An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 2019 Apr ;15(4) :e1006650. 62
- Boulet N, Boussere A, Mezzarobba M, Sofonea MT, Payen D, Lipman J, et al. Intensive Care Unit activity in France from the national database between 2013 and 2019 : More critically ill patients, shorter stay and lower mortality rate. *Anaesthesia Critical Care & Pain Medicine* 2023 Oct ;42(5) :101228. 38
- Britton T, Scalia Tomba G. Estimation in emerging epidemics : biases and remedies. *Journal of The Royal Society Interface* 2019 Jan ;16(150) :20180670. 41, 63, 72
- Byambasuren O, Cardona M, Bell K, Clark J, McLaws ML, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission : Systematic review and meta-analysis. *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada* 2020 Dec ;5(4) :223–234. 44
- Callaway E. Why a highly mutated coronavirus variant has scientists on alert. *Nature* 2023 Aug ;620(7976) :934–934. 61
- Calvez V, Korobeinikov A, Maini PK. Cluster formation for multi-strain infections with cross-immunity. *Journal of Theoretical Biology* 2005 Mar ;233(1) :75–83. 61
- CDC. Lassa Fever. Centers for Disease Control and Prevention; 2019. 70
- Cohen JE. Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better. *PLOS Biology* 2004 Dec ;2(12) :e439. 38
- Cooper BS, Medley GF, Bradley SJ, Scott GM. An Augmented Data Method for the Analysis of Nosocomial Infection Data. *American Journal of Epidemiology* 2008 Sep ;168(5) :548–557. 72
- Cox DR, Miller HD. *The theory of stochastic processes*, vol. 134. CRC press; 1977. 45
- Currie CSM, Fowler JW, Kotiadis K, Monks T, Onggo BS, Robertson DA, et al. How simulation modelling can help reduce the impact of COVID-19. *Journal of Simulation* 2020 Apr ;14(2) :83–97. 38
- Danesh G, Elie B, Michalakis Y, Sofonea MT, Bal A, Behillil S, et al. Early phylodynamics analysis of the COVID-19 epidemic in France. *Peer Community Journal* 2021 ;1. 57, 61

- Danesh G, Saulnier E, Gascuel O, Choisy M, Alizon S. TiPS : Rapidly simulating trajectories and phylogenies from compartmental models. *Methods in Ecology and Evolution* 2023 ;14(2) :487–495. [62](#)
- Davies NG, Jarvis CI, Edmunds WJ, Jewell NP, Diaz-Ordaz K, Keogh RH. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* 2021 Mar ;p. 1–5. [46](#)
- Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID : major findings, mechanisms and recommendations. *Nature Reviews Microbiology* 2023 Mar ;21(3) :133–146. [59](#), [66](#)
- Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the basic reproduction number (R0). *Emerging infectious diseases* 2019 ;25(1) :1. [71](#)
- Dellicour S, Lequime S, Vrancken B, Gill MS, Bastide P, Gangavarapu K, et al. Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework. *Nature Communications* 2020 Nov ;11(1) :5620. [62](#)
- Delussu F, Tizzoni M, Gauvin L. Evidence of pandemic fatigue associated with stricter tiered COVID-19 restrictions. *PLOS Digital Health* 2022 May ;1(5) :e0000035. [50](#), [60](#), [63](#)
- Dennis JM, McGovern AP, Vollmer SJ, Mateen BA. Improving Survival of Critical Care Patients With Coronavirus Disease 2019 in England : A National Cohort Study, March to June 2020\*. *Critical Care Medicine* 2021 Feb ;49(2) :209–214. [46](#)
- Destoumieux-Garzón D, Mavingui P, Boetsch G, Boissier J, Darriet F, Duboz P, et al. The One Health Concept : 10 Years Old and a Long Road Ahead. *Frontiers in Veterinary Science* 2018 ;5. [67](#), [74](#)
- Diekmann O, Heesterbeek H, Britton T. Mathematical tools for understanding infectious disease dynamics, vol. 7. Princeton University Press ; 2013. [39](#), [41](#)
- Diekmann O, Inaba H. A systematic procedure for incorporating separable static heterogeneity into compartmental epidemic models. *Journal of Mathematical Biology* 2023 Jan ;86(2) :29. [48](#)
- Diekmann O, Othmer HG, Planqué R, Bootsma MCJ. The discrete-time Kermack–McKendrick model : A versatile and computationally attractive framework for modeling epidemics. *Proceedings of the National Academy of Sciences* 2021 Sep ;118(39) :e2106332118. [47](#), [48](#)
- Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A. The e1071 Package. *R package* 2006 Feb ;1 :5–24. [74](#)
- Djidjou-Demasse R, Michalakis Y, Choisy M, Sofonea MT, Alizon S. Optimal COVID-19 epidemic control until vaccine deployment. *Infectious Diseases (except HIV/AIDS)* ; 2020. [56](#)
- Djidjou-Demasse R, Alizon S, Sofonea MT. Within-host bacterial growth dynamics with both mutation and horizontal gene transfer. *Journal of Mathematical Biology* 2021 Feb ;82(3) :16. [38](#)
- Djidjou-Demasse R, Selinger C, Sofonea MT. Épidémiologie mathématique et modélisation de la pandémie de Covid-19 : enjeux et diversité. *Revue Francophone des Laboratoires* 2020 Nov ;2020(526) :63–69. [38](#)
- Djidjou-Demasse R, Sofonea MT, Choisy M, Alizon S. Within-host evolutionary dynamics of antimicrobial quantitative resistance. *Mathematical Modelling of Natural Phenomena* 2023 Jun ;. [38](#)
- ECDC. Risk assessment : Monkeypox multi-country outbreak. European Centre for Disease Prevention and Control ; 2022. [66](#), [70](#)
- Elie B, Roquebert B, Sofonea MT, Trombert-Paolantoni S, Foulongne V, Guedj J, et al. Variant-specific SARS-CoV-2 within-host kinetics. *Journal of Medical Virology* 2022 ;94(8) :3625–3633. [56](#)
- Elie B, Selinger C, Alizon S. The source of individual heterogeneity shapes infectious disease outbreaks. *Proceedings of the Royal Society B : Biological Sciences* 2022 May ;289(1974) :20220232. [41](#)
- Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9 : Impact of non-pharmaceutical interventions (NPIs) to reduce COVID- 19 mortality and healthcare demand. [imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/covid-19-reports/](https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/covid-19-reports/) ; 2020. [38](#), [49](#), [50](#)

- Ferkol T, Schraufnagel D. The Global Burden of Respiratory Disease. *Annals of the American Thoracic Society* 2014 Mar;11(3) :404–406. [60](#)
- Ferretti L, Ledda A, Wymant C, Zhao L, Ledda V, Abeler L, et al. The timing of COVID-19 transmission. *medRxiv* 2020 Sep; : [44](#)
- Fitz-Simon N, Ferguson J, Alvarez-Iglesias A, Sofonea MT, Kamiya T. Understanding the role of mask-wearing during COVID-19 on the island of Ireland. *Royal Society Open Science* 2023 Jul;10(7) :221540. [82](#)
- Flahault A, Calmy A, Costagliola D, Drapkina O, Eckerle I, Larson HJ, et al. No time for complacency on COVID-19 in Europe. *The Lancet* 2023 Jun;401(10392) :1909–1912. [59](#), [61](#)
- Foppa IM. A historical introduction to mathematical modeling of infectious diseases : Seminal Papers in Epidemiology. Academic Press; 2016. [41](#)
- Forien R, Pang G, Pardoux E. Estimating the state of the Covid-19 epidemic in France using a non-Markovian model. *medRxiv* 2020 Jul;p. 2020.06.27.20141671. [51](#)
- Fraser C, Riley S, Anderson RM, Ferguson NM. Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences* 2004 Apr;101(16) :6146–6151. [69](#)
- Fährnich C, Denecke K, Adeoye OO, Benzler J, Claus H, Kirchner G, et al. Surveillance and Outbreak Response Management System (SORMAS) to support the control of the Ebola virus disease outbreak in West Africa. *Eurosurveillance* 2015 Mar;20(12) :21071. [64](#)
- Gallet R, Fabre F, Thébaud G, Sofonea MT, Sicard A, Blanc S, et al. Small Bottleneck Size in a Highly Multipartite Virus during a Complete Infection Cycle. *Journal of Virology* 2018 May;92(14) :e00139–18, /jvi/92/14/e00139–18.atom. [38](#)
- Ganser I, Buckeridge DL, Heffernan JM, Prague M, Thiébaud R. Estimating the Population Effectiveness of Interventions Against COVID-19 in France : A Modelling Study. *Available at SSRN 4488069* 2023 Jun; : [46](#)
- Gaudart J, Landier J, Huiart L, Legendre E, Lehot L, Bendiane MK, et al. Factors associated with the spatial heterogeneity of the first wave of COVID-19 in France : a nationwide geo-epidemiological study. *The Lancet Public Health* 2021 Apr;6(4) :e222–e231. [41](#), [60](#), [62](#)
- Gaudou B, Huynh NQ, Philippon D, Brugière A, Chapuis K, Taillandier P, et al. COMOKIT : A Modeling Kit to Understand, Analyze, and Compare the Impacts of Mitigation Policies Against the COVID-19 Epidemic at the Scale of a City. *Frontiers in Public Health* 2020;8. [39](#)
- Gayat E, Raux M. From descriptive epidemiology to interventional epidemiology : The central role of epidemiologists in COVID-19 crisis management. *Anaesthesia Critical Care & Pain Medicine* 2022 Apr;41(2) :101056. [59](#)
- Gaynard A, Bosetti P, Feri A, Destras G, Enouf V, Andronico A, et al. Early assessment of diffusion and possible expansion of SARS-CoV-2 Lineage 20I/501Y.V1 (B.1.1.7, variant of concern 202012/01) in France, January to March 2021. *Eurosurveillance* 2021 Mar;26(9) :2100133. [46](#)
- Geismar C, Nguyen V, Fragaszy E, Shrotri M, Navaratnam AMD, Beale S, et al. Bayesian reconstruction of SARS-CoV-2 transmissions highlights substantial proportion of negative serial intervals. *Epidemics* 2023 Sep;44 :100713. [41](#), [44](#), [63](#)
- Goldberg Y, Mandel M, Bar-On YM, Bodenheimer O, Freedman LS, Ash N, et al. Protection and Waning of Natural and Hybrid Immunity to SARS-CoV-2. *New England Journal of Medicine* 2022 Jun;386(23) :2201–2212. [60](#), [63](#)
- Goldstein RE. Are theoretical results ‘Results’? *eLife* 2018 Jul;7 :e40018. [39](#)
- Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. The species Severe acute respiratory syndrome-related coronavirus : classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* 2020 Apr;5(4) :536–544. [38](#)

- Guerra FM, Bolotin S, Lim G, Heffernan J, Deeks SL, Li Y, et al. The basic reproduction number (R0) of measles : a systematic review. *The Lancet Infectious Diseases* 2017 Dec;17(12) :e420–e428. **71**
- Hafenbrädl S, Waeger D, Marewski JN, Gigerenzer G. Applied Decision Making With Fast-and-Frugal Heuristics. *Journal of Applied Research in Memory and Cognition* 2016 Jun;5(2) :215–231. **76**
- Haim-Boukobza S, Roquebert B, Trombert-Paolantoni S, Lecorche E, Verdurme L, Foulongne V, et al. Detecting Rapid Spread of SARS-CoV-2 Variants, France, January 26-February 16, 2021. *Emerging Infectious Diseases* 2021;27(5) :10. **57**
- Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. *Nature Human Behaviour* 2020 Nov;p. 1–10. **38, 63**
- Hay JA, Kennedy-Shaffer L, Kanjilal S, Lennon NJ, Gabriel SB, Lipsitch M, et al. Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science* 2021 Jun;. **56**
- He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine* 2020 May;26(5) :672–675. **63, 69**
- Hellewell J, Russell TW, Matthews R, Severn A, Adam S, Enfield L, et al. Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. *BMC Medicine* 2021 Apr;19(1) :106. **44**
- Hillis SD, Unwin HJT, Chen Y, Cluver L, Sherr L, Goldman PS, et al. Global minimum estimates of children affected by COVID-19-associated orphanhood and deaths of caregivers : a modelling study. *The Lancet* 2021 Jul;398(10298) :391–402. **59**
- Hoertel N, Blachier M, Blanco C, Olfson M, Massetti M, Rico MS, et al. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature Medicine* 2020 Jul;p. 1–5. **51**
- Hyndman RJ, Athanasopoulos G. Forecasting : principles and practice. OTexts; 2018. **72**
- Imai N, Dorigatti I, Cori A, Riley S, Ferguson N. Estimating the potential total number of novel Coronavirus cases in Wuhan City, China. [imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/covid-19-reports/](https://imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/covid-19-reports/); 2020. **38, 60**
- Ipsos. Baromètre "Science et Société" - Vague 2; 2022. **66**
- Josse J, Husson F. missMDA : A Package for Handling Missing Values in Multivariate Data Analysis. *Journal of Statistical Software* 2016 Apr;70 :1–31. **72**
- Kamiya T, Alvarez-Iglesias A, Ferguson J, Murphy S, Sofonea MT, Fitz-Simon N. Estimating time-dependent contact : a multi-strain epidemiological model of SARS-CoV-2 on the island of Ireland. *Global Epidemiology* 2023 Dec;5 :100111. **51, 82**
- Karadag E. Increase in COVID-19 cases and case-fatality and case-recovery rates in Europe : A cross-temporal meta-analysis. *Journal of Medical Virology* 2020;92(9) :1511–1517. **72**
- Karatayev VA, Anand M, Bauch CT. Local lockdowns outperform global lockdown on the far side of the COVID-19 epidemic curve. *Proceedings of the National Academy of Sciences* 2020 Sep;p. 202014385. **50, 59, 60**
- Keeling MJ. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London Series B : Biological Sciences* 1999 Apr;266(1421) :859–867. **62**
- Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2008. **39, 41**
- Keeling MJ, Rohani P. Estimating spatial coupling in epidemiological systems : a mechanistic approach. *Ecology Letters* 2002;5(1) :20–29. **62**
- Kermack W, McKendrick A. Contributions to the mathematical theory of epidemics, part 1. *Proceedings of the Royal Society A : Mathematical, Physical and Engineering Science* 1927;. **41**

- von Kügelgen J, Gresele L, Schölkopf B. Simpson's Paradox in COVID-19 Case Fatality Rates : A Mediation Analysis of Age-Related Causal Effects. *IEEE Transactions on Artificial Intelligence* 2021 Feb ;2(1) :18–27. [41](#)
- Lakdawala SS, Menachery VD. Catch me if you can : superspreading of COVID-19. *Trends in Microbiology* 2021 May ;p. S0966842X21001244. [62](#), [63](#)
- Le Cann P, Batéjat C, Delache B, Garcia-Verdugo I, Ha TL, Leclercq I, et al. Note d'appui scientifique et technique de l'Anses relatif à la viabilité dans l'air et la dose infectante du virus SARS-COV-2. Anses ; 2021. [61](#)
- Lefrant JY, Pirracchio R, Benhamou D, Dureuil B, Pottecher J, Samain E, et al. ICU bed capacity during COVID-19 pandemic in France : From ephemeral beds to continuous and permanent adaptation. *Anaesthesia Critical Care & Pain Medicine* 2021 Jun ;40(3) :100873. [43](#)
- Lewis AC, Jenkins D, Whitty CJM. Hidden harms of indoor air pollution — five steps to expose them. *Nature* 2023 Feb ;614(7947) :220–223. [59](#)
- Lewis D. Indoor air is full of flu and COVID viruses. Will countries clean it up? *Nature* 2023 Mar ;615(7951) :206–208. [59](#)
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus - Infected Pneumonia. *New England Journal of Medicine* 2020 Mar ;382(13) :1199–1207. [38](#)
- Lind ML, Dorion M, Houde AJ, Lansing M, Lapidus S, Thomas R, et al. Evidence of leaky protection following COVID-19 vaccination and SARS-CoV-2 infection in an incarcerated population. *Nature Communications* 2023 Aug ;14(1) :5055. [60](#), [63](#)
- Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung Sm, et al. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation : A Statistical Analysis of Publicly Available Case Data. *Journal of Clinical Medicine* 2020 Feb ;9(2) :538. [44](#)
- Lipsitch M, Donnelly CA, Fraser C, Blake IM, Cori A, Dorigatti I, et al. Potential Biases in Estimating Absolute and Relative Case-Fatality Risks during Outbreaks. *PLOS Neglected Tropical Diseases* 2015 Jul ;9(7) :e0003846. [41](#)
- Liu Q, Qin C, Liu M, Liu J. Effectiveness and safety of SARS-CoV-2 vaccine in real-world studies : a systematic review and meta-analysis. *Infectious Diseases of Poverty* 2021 Nov ;10(1) :132. [46](#)
- Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. *Nature* 2005 Nov ;438(7066) :355–359. [41](#), [51](#)
- Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of SARS-CoV-2. *Nature Reviews Microbiology* 2023 Apr ;p. 1–19. [60](#), [61](#)
- Markov PV, Katzourakis A, Stilianakis NI. Antigenic evolution will lead to new SARS-CoV-2 variants with unpredictable severity. *Nature Reviews Microbiology* 2022 Mar ;p. 1–2. [61](#)
- Massey A, Boennec C, Restrepo-Ortiz CX, Blanchet C, Alizon S, Sofonea MT. Real-time forecasting of COVID-19-related hospital strain in France using a non-Markovian mechanistic model. *medRxiv* 2023 Feb ;. [33](#), [54](#), [59](#), [61](#)
- McAloon C, Collins A, Hunt K, Barber A, Byrne AW, Butler F, et al. Incubation period of COVID-19 : a rapid systematic review and meta-analysis of observational research. *BMJ Open* 2020 Aug ;10(8) :e039652. [44](#), [73](#)
- McAndrew T, Majumder MS, Lover AA, Venkatramanan S, Bocchini P, Besiroglu T, et al. Early human judgment forecasts of human monkeypox, May 2022. *The Lancet Digital Health* 2022 Aug ;4(8) :e569–e571. [71](#)
- Meganck RM, Baric RS. Developing therapeutic approaches for twenty-first-century emerging infectious viral diseases. *Nature Medicine* 2021 Mar ;27(3) :401–410. [66](#), [67](#)
- Mercer TR, Salit M. Testing at scale during the COVID-19 pandemic. *Nature Reviews Genetics* 2021 May ;p. 1–12. [44](#)

- Michalakis Y, Sofonea MT, Alizon S, Bravo IG. SARS-CoV-2 viral RNA levels are not 'viral load'. *Trends in Microbiology* 2021 Nov;29(11) :970–972. 57
- Miller JC. A Note on the Derivation of Epidemic Final Sizes. *Bulletin of Mathematical Biology* 2012 Sep;74(9) :2125–2141. 51, 75
- Monge S, Pastor-Barriuso R, Hernán MA. The imprinting effect of covid-19 vaccines : an expected selection bias in observational studies. *BMJ* 2023 Jun;381 :e074404. 41
- Mora C, McKenzie T, Gaw IM, Dean JM, von Hammerstein H, Knudson TA, et al. Over half of known human pathogenic diseases can be aggravated by climate change. *Nature Climate Change* 2022 Aug;p. 1–7. 59
- Msemburi W, Karlinsky A, Knutson V, Aleshin-Guendel S, Chatterji S, Wakefield J. The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature* 2023 Jan;613(7942) :130–137. 59
- Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. *Nature Medicine* 2021 Mar;. 44
- Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *International Journal of Infectious Diseases* 2020 Apr;93 :284–286. 44
- Noble WS. What is a support vector machine? *Nature Biotechnology* 2006 Dec;24(12) :1565–1567. 76
- North AR, Godfray HCJ. The dynamics of disease in a metapopulation : The role of dispersal range. *Journal of Theoretical Biology* 2017 Apr;418 :57–65. 62
- Nyberg T, Ferguson NM, Nash SG, Webster HH, Flaxman S, Andrews N, et al. 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. *The Lancet* 2022 Apr;399(10332) :1303–1312. 46
- Oberndorfer M, Dorner TE, Brunnmayr M, Berger K, Dugandzic B, Bach M. Health-related and socio-economic burden of the COVID-19 pandemic in Vienna. *Health & Social Care in the Community* 2022 ;30(4) :1550–1561. 66
- O'Driscoll M, Harry C, Donnelly CA, Cori A, Dorigatti I. A comparative analysis of statistical methods to estimate the reproduction number in emerging epidemics with implications for the current COVID-19 pandemic. *Epidemiology*; 2020. 46
- O'Hara-Wild M, Hyndman R, Wang E, Caceres G, Hensel T, Hyndman T, fable - Forecasting models for tidy time series. CRAN; 2020. <https://github.com/tidyverts/fable>. 55
- Ovaskainen O, Cornell SJ. Space and stochasticity in population dynamics. *Proceedings of the National Academy of Sciences* 2006;103(34) :12781–12786. 62
- Ovaskainen O, Finkelshtein D, Kutoviy O, Cornell S, Bolker B, Kondratiev Y. A general mathematical framework for the analysis of spatiotemporal point processes. *Theoretical ecology* 2014;7(1) :101–113. 62
- Paireau J, Andronico A, Hozé N, Layan M, Crépey P, Roumagnac A, et al. An ensemble model based on early predictors to forecast COVID-19 health care demand in France. *Proceedings of the National Academy of Sciences* 2022 May;119(18) :e2103302119. 38, 42, 59
- Pangallo M, Aleta A, Chanona RMdR, Pichler A, Martín-Corral D, Chinazzi M, et al. The unequal effects of the health-economy tradeoff during the COVID-19 pandemic. *arXiv : 221203567* 2022 Dec;. 50, 59
- Peng Z, Rojas ALP, Kropff E, Bahnfleth W, Buonanno G, Dancer SJ, et al. Practical Indicators for Risk of Airborne Transmission in Shared Indoor Environments and Their Application to COVID-19 Outbreaks. *Environmental Science & Technology* 2022 Jan;p. acs.est.1c06531. 60, 63
- Perry BL, Aronson B, Pescosolido BA. Pandemic precarity : COVID-19 is exposing and exacerbating inequalities in the American heartland. *Proceedings of the National Academy of Sciences* 2021 Feb;118(8) :e2020685118. 66

- Phillips ND, Neth H, Woike JK, Gaissmaier W. FFTrees : A toolbox to create, visualize, and evaluate fast-and-frugal decision trees. *Judgment and Decision Making* 2017 Jul;12(4) :344–368. [74](#)
- Polak SB, Van Gool IC, Cohen D, von der Thüsen JH, van Paassen J. A systematic review of pathological findings in COVID-19 : a pathophysiological timeline and possible mechanisms of disease progression. *Modern Pathology* 2020 Nov;33(11) :2128–2138. [44](#)
- Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health* 2021 Apr;20 :100178. [76](#)
- Prunas O, Warren JL, Crawford FW, Gazit S, Patalon T, Weinberger DM, et al. Vaccination with BNT162b2 reduces transmission of SARS-CoV-2 to household contacts in Israel. *Science* 2022 Jan;. [52](#)
- Public Health England. SARS-CoV-2 variants of concern and variants under investigation - Technical briefing 15; 2021. [51](#)
- Pulliam JRC, van Schalkwyk C, Govender N, von Gottberg A, Cohen C, Groome MJ, et al. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science* 2022 Mar;376(6593) :eabn4947. [60](#)
- Puranik A, Lenahan PJ, Silvert E, Niesen MJM, Corchado-Garcia J, O'Horo JC, et al. Comparison of two highly-effective mRNA vaccines for COVID-19 during periods of Alpha and Delta variant prevalence. *Public and Global Health*; 2021. [51](#)
- Randles J. The 'pathosphere', paradigms and enigmatic pathogens. *Australasian Plant Pathology : AAP* 1999;28(4) :263. [67](#)
- Reyné B, Danesh G, Alizon S, Sofonea MT. Rt2 : computing and visualising COVID-19 epidemics temporal reproduction number. *medRxiv* 2020 Dec;. [30](#)
- Reyné B, Richard Q, Selinger C, Sofonea MT, Djidjou-Demasse R, Alizon S. Non-Markovian modelling highlights the importance of age structure on Covid-19 epidemiological dynamics. *Mathematical Modelling of Natural Phenomena* 2022;17 :7. [48](#)
- Reyné B, Selinger C, Sofonea MT, Miot S, Pisoni A, Tuaille E, et al. Analysing different exposures identifies that wearing masks and establishing COVID-19 areas reduce secondary-attack risk in aged-care facilities. *International Journal of Epidemiology* 2021 Jun;p. dyab121. [56](#)
- Reyné B, Sofonea MT, Alizon S. Modelling long-term COVID-19 hospital admission dynamics using immune protection waning data. *medRxiv*; 2022. [30](#)
- Richard Q, Alizon S, Choisy M, Sofonea MT, Djidjou-Demasse R. Age-structured non-pharmaceutical interventions for optimal control of COVID-19 epidemic. *PLOS Computational Biology* 2021 Mar;17(3) :e1008776. [56](#), [61](#)
- Roquebert B, Haim-Boukoba S, Trombert-Paolantoni S, Lecorche E, Verdurme L, Foulongne V, et al. SARS-CoV-2 variants of concern are associated with lower RT-PCR amplification cycles between January and March 2021 in France. *International Journal of Infectious Diseases* 2021 Dec;113 :12–14. [56](#), [57](#), [61](#)
- Roquebert B, Trombert-Paolantoni S, Haim-Boukoba S, Lecorche E, Verdurme L, Foulongne V, et al. The SARS-CoV-2 B.1.351 lineage (VOC Beta) is outgrowing the B.1.1.7 lineage (VOC Alpha) in some French regions in April 2021. *Eurosurveillance* 2021 Jun;26(23) :2100447. [57](#)
- Roques L, Klein EK, Papaix J, Sar A, Soubeyrand S. Impact of Lockdown on the Epidemic Dynamics of COVID-19 in France. *Frontiers in Medicine* 2020;7. [51](#)
- Rothman KJ, Greenland S. Causation and Causal Inference in Epidemiology. *American Journal of Public Health* 2005 Jul;95(S1) :S144–S150. [39](#)
- Roucaute D. La triple épidémie de Covid-19, de grippe et de bronchiolite reflue en France, mais les tensions perdurent dans les hôpitaux. *Le Monde* 2023 Jan;. [60](#), [71](#)

- Roy M, Pascual M. On representing network heterogeneities in the incidence rate of simple epidemic models. *Ecological Complexity* 2006 Mar;3(1) :80–90. **62**
- Russell TW, Hellewell J, Jarvis CI, van Zandvoort K, Abbott S, Ratnayake R, et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. *Eurosurveillance* 2020 Mar;25(12). **41, 46**
- 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 May;p. eabc3517. **51**
- Santé publique France, Données hospitalières relatives à l'épidémie de COVID-19 (SI-VIC) - data.gouv.fr; 2020. <https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/>. **48, 58**
- Santé publique France, Données relatives aux résultats des tests virologiques COVID-19 (SI-DEP) - data.gouv.fr; 2020. <https://www.data.gouv.fr/fr/datasets/old-donnees-relatives-aux-resultats-des-tests-virologiques-covid-19/>. **58**
- Santé Publique France, Données relatives aux personnes vaccinées contre la Covid-19 (VAC-SI) - data.gouv.fr; 2021. <https://www.data.gouv.fr/fr/datasets/donnees-relatives-aux-personnes-vaccinees-contre-la-covid-19-1/>. **46**
- Santé publique France. Bulletin épidémiologique bronchiolite. Bilan de la surveillance 2022-2023.; 2023. **60**
- Sasaki A, Lion S, Boots M. Antigenic escape selects for the evolution of higher pathogen transmission and virulence. *Nature Ecology & Evolution* 2022 Jan;6(1) :51–62. **61**
- Saulnier E, Gascuel O, Alizon S. Inferring epidemiological parameters from phylogenies using regression-ABC : A comparative study. *PLOS Computational Biology* 2017 Mar;13(3) :e1005416. **62**
- Semenzato L, Botton J, Drouin J, Cuenot F, Dray-Spira R, Weill A, et al. Chronic diseases, health conditions and risk of COVID-19-related hospitalization and in-hospital mortality during the first wave of the epidemic in France : a cohort study of 66 million people. *The Lancet Regional Health – Europe* 2021 Sep;8. **46**
- Servedio MR, Brandvain Y, Dhole S, Fitzpatrick CL, Goldberg EE, Stern CA, et al. Not just a theory—the utility of mathematical models in evolutionary biology. *PLOS Biology* 2014;12(12) :e1002017. **38**
- Siegenfeld AF, Taleb NN, Bar-Yam Y. Opinion : What models can and cannot tell us about COVID-19. *Proceedings of the National Academy of Sciences* 2020 Jul;117(28) :16092–16095. **62**
- Simpson S, Kaufmann MC, Glozman V, Chakrabarti A. Disease X : accelerating the development of medical countermeasures for the next pandemic. *The Lancet Infectious Diseases* 2020 May;20(5) :e108–e115. **66, 67**
- Sofonea MT. A global asymptotic stability condition for a Lotka–Volterra model with indirect interactions. *Applicable Analysis* 2019 Jul;98(9) :1636–1645. **38**
- Sofonea MT, Alizon S. Anticipating COVID-19 intensive care unit capacity strain : A look back at epidemiological projections in France. *Anaesthesia, Critical Care & Pain Medicine* 2021 Aug;40(4) :100943. **39, 54, 61, 64**
- Sofonea MT, Alizon S, Michalakis Y. From within-host interactions to epidemiological competition : a general model for multiple infections. *Philosophical Transactions of the Royal Society B : Biological Sciences* 2015 Aug;370(1675) :20140303. **61, 64, 65**
- Sofonea MT, Alizon S, Michalakis Y. Exposing the diversity of multiple infection patterns. *Journal of Theoretical Biology* 2017;419 :278 – 289. **61**
- Sofonea MT, Boennec C, Fontalirant A, Regnat M, Lefrant JY, Alizon S, et al. Quantifying the real-life impacts of vaccination on critical COVID-19. OSF Preprints; 2021. **33, 52**
- Sofonea MT, Boennec C, Michalakis Y, Alizon S. Two waves and a high tide : the COVID-19 epidemic in France. *Anaesthesia Critical Care & Pain Medicine* 2021 Jun;40(3) :100881. **53**



- Sofonea MT, Cauchemez S, Boëlle PY. Epidemic models : why and how to use them. *Anaesthesia Critical Care & Pain Medicine* 2022 Apr ;41(2) :101048. **38, 60**
- Sofonea MT, Reyné B, Alizon S, COVIDSIM-FR - Combining statistical analysis of hospital data and parsimonious non Markovian modelling for inferring epidemiological parameters and simulating NPI of the COVID-19 epidemic in France. *i-Trop bioinformatics* ; 2020. <https://bioinfo-shiny.ird.fr/COVIDSIM2-fr/>. **30, 53, 82**
- Sofonea MT, Reyné B, Elie B, Djidjou-Demasse R, Selinger C, Michalakis Y, et al. Epidemiological monitoring and control perspectives : application of a parsimonious modelling framework to the COVID-19 dynamics in France. *medRxiv* 2020 May ;p. 2020.05.22.20110593. **30, 49, 82**
- Sofonea MT, Reyné B, Elie B, Djidjou-Demasse R, Selinger C, Michalakis Y, et al. Memory is key in capturing COVID-19 epidemiological dynamics. *Epidemics* 2021 Jun ;35 :100459. **39, 45, 46, 47, 49, 51, 61, 64, 70**
- Sofonea MT, Roquebert B, Foulongne V, Morquin D, Verdurme L, Trombert-Paolantoni S, 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* 2022 Jul ;28(7) :1355–1365. **39, 46, 54, 57, 60, 61**
- Sonabend R, Whittles LK, Imai N, Perez-Guzman PN, Knock ES, Rawson T, et al. Non-pharmaceutical interventions, vaccination, and the SARS-CoV-2 delta variant in England : a mathematical modelling study. *The Lancet* 2021 Nov ;398(10313) :1825–1835. **60**
- Sportisse A, Boyer C, Josse J. Imputation and low-rank estimation with Missing Not At Random data. *Statistics and Computing* 2020 Nov ;30(6) :1629–1643. **72**
- Stowe J, Tessier E, Zhao H, Guy R, Muller-Pebody B, Zambon M, et al. Interactions between SARS-CoV-2 and influenza, and the impact of coinfection on disease severity : a test-negative design. *International Journal of Epidemiology* 2021 Aug ;50(4) :1124–1133. **60**
- Supplisson O, Sofonea MT. Pièges et mésusages en analyse de données. *Anaesthesia, Critical Care & Pain Medicine* 2023 ;(in press). **41**
- Surkova E, Nikolayevskyy V, Drobniewski F. False-positive COVID-19 results : hidden problems and costs. *The Lancet Respiratory Medicine* 2020 Dec ;8(12) :1167–1168. **41**
- Swets MC, Russell CD, Harrison EM, Docherty AB, Lone N, Girvan M, et al. SARS-CoV-2 co-infection with influenza viruses, respiratory syncytial virus, or adenoviruses. *The Lancet* 2022 Apr ;399(10334) :1463–1464. **60**
- Tan ST, Kwan AT, Rodríguez-Barraquer I, Singer BJ, Park HJ, Lewnard JA, et al. Infectiousness of SARS-CoV-2 breakthrough infections and reinfections during the Omicron wave. *Nature Medicine* 2023 Feb ;29(2) :358–365. **52**
- Tan W, Zhao X, Ma X, Wang W, Niu P, Xu W, et al. A Novel Coronavirus Genome Identified in a Cluster of Pneumonia Cases - Wuhan, China 2019-2020. *China CDC Weekly* 2020 Jan ;2(4) :61–62. **38**
- Tartof SY, Slezak JM, Puzniak L, Hong V, Frankland TB, Ackerson BK, et al. BNT162b2 vaccine effectiveness against SARS-CoV-2 omicron BA.4 and BA.5. *The Lancet Infectious Diseases* 2022 Dec ;22(12) :1663–1665. **60**
- Thibert C. Covid-19 : un confinement le week-end est-il une mesure assez forte pour vider les hôpitaux ? *Le Figaro* 2021 Mar ;. **49**
- Thomine O, Alizon S, Boennec C, Barthelemy M, Sofonea M. Emerging dynamics from high-resolution spatial numerical epidemics. *eLife* 2021 Oct ;10 :e71417. **42, 51, 59, 61, 62, 64**
- Tran Kiem C, Massonnaud CR, Levy-Bruhl D, Poletto C, Colizza V, Bosetti P, et al. A modelling study investigating short and medium-term challenges for COVID-19 vaccination : From prioritisation to the relaxation of measures. *eClinicalMedicine* 2021 Aug ;38 :101001. **38**

- Trapman P, Ball F, Dhersin JS, Tran VC, Wallinga J, Britton T. Inferring  $R_0$  in emerging epidemics—the effect of common population structure is small. *Journal of The Royal Society Interface* 2016 Aug;13(121) :20160288. 41, 51
- Twohig KA, Nyberg T, Zaidi A, Thelwall S, Sinnathamby MA, Aliabadi S, et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern : a cohort study. *The Lancet Infectious Diseases* 2022 Jan;22(1) :35–42. 46, 61
- UKHSA. COVID-19 vaccine surveillance report : week 23. UK Health Security Agency; 2023. 60
- Van Kerkhove MD, Ryan MJ, Ghebreyesus TA. Preparing for “Disease X”. *Science* 2021 Oct;374(6566) :377–377. 66
- Velly L, Gayat E, Quintard H, Weiss E, De Jong A, Cuvillon P, et al. Guidelines : Anaesthesia in the context of COVID-19 pandemic. *Anaesthesia Critical Care & Pain Medicine* 2020 Jun;39(3) :395–415. 46
- Verdume L, Danesh G, Trombert S, Sofonea MT, Noel V, Foulongne V, et al. Spread of SARS-CoV-2 Delta variant infections bearing the S :E484Q and S :T95I mutations in July and August 2021 in France; 2021. 57
- Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019 : a model-based analysis. *The Lancet Infectious Diseases* 2020 Mar;p. S1473309920302437. 46
- Villanueva DDH, Arcega V, Rao M. Review of respiratory syncytial virus infection among older adults and transplant recipients. *Therapeutic Advances in Infectious Disease* 2022 Jan;9 :20499361221091413. 60
- Volz E. Fitness, growth and transmissibility of SARS-CoV-2 genetic variants. *Nature Reviews Genetics* 2023 Jun ; 61, 63
- 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 May;593(7858) :266–269. 46
- Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019 : a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 2020 Oct;396(10258) :1204–1222. 60
- Wang H, Paulson KR, Pease SA, Watson S, Comfort H, Zheng P, et al. Estimating excess mortality due to the COVID-19 pandemic : a systematic analysis of COVID-19-related mortality, 2020-21. *The Lancet* 2022 Apr;399(10334) :1513–1536. 59, 66
- 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 Jun;0(0). 52, 53
- WHO. List of Blueprint priority diseases. World Health Organization (Geneva); 2018. 66, 69
- WHO. The top 10 causes of death. World Health Organization (Geneva); 2020. 60
- WHO. COVID-19 Dashboard. World Health Organization (Geneva); 2023. 66
- WHO. WHO Fact sheets - Ebola virus disease. World Health Organization (Geneva); 2023. 70
- Wilks SS. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics* 1938 ;9(1) :60–62. 46
- Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. *Nature* 2007 May;447(7142) :279–283. 71
- Woolhouse MEJ, Brierley L. Epidemiological characteristics of human-infective RNA viruses. *Scientific Data* 2018 Feb;5(1) :180017. 67
- Woolhouse MEJ, Brierley L, McCaffery C, Lycett S. Assessing the Epidemic Potential of RNA and DNA Viruses. *Emerging Infectious Diseases* 2016 Dec;22(12) :2037–2044. 67

- Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020 Mar;579(7798) :265–269. 38
- Xie Y, Xu E, Bowe B, Al-Aly Z. Long-term cardiovascular outcomes of COVID-19. *Nature Medicine* 2022 Mar;28(3) :583–590. 59
- Yang Q, Saldi TK, Gonzales PK, Lasda E, Decker CJ, Tat KL, et al. Just 2% of SARS-CoV-2-positive individuals carry 90% of the virus circulating in communities. *Proceedings of the National Academy of Sciences* 2021 May;118(21) :e2104547118. 60
- Zelner J, Riou J, Etzioni R, Gelman A. Accounting for uncertainty during a pandemic. *Patterns* 2021 Aug;2(8). 41, 60, 74
- Zhang Y, Zhang A, Wang J. Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China. *Transport Policy* 2020 Aug;94 :34–42. 67
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* 2020 Feb;382(8) :727–733. 38
- Zwart MP, Blanc S, Johnson M, Manrubia S, Michalakis Y, Sofonea MT. Unresolved advantages of multipartitism in spatially structured environments. *Virus Evolution* 2021 Jan;p. veab004. 38
- Équipes de surveillance de la grippe (Santé publique France). Surveillance de la grippe en France, saison 2021-2022. *Bulletin épidémiologique hebdomadaire* 2022;21 :362–75. 60

## 4.3 Remerciements

En premier lieu, j'exprime toute ma gratitude envers les Professeur(e)s Antoine Flahault, Mahmoud Zureik, Vincent Calvez, Jérémie Sellam – pour avoir accepté de rapporter mon dossier d'Habilitation –, Solen Kernéis, Vittoria Colizza et Dominique Costagliola – pour avoir accepté d'en examiner la soutenance. Je mesure l'immense honneur qu'ils et elles me font en allouant un peu de leur précieux temps aux travaux ici présentés.

Ces travaux, et la raison même de prétendre à l'Habilitation, n'auraient vu le jour sans les membres qui ont composé la merveilleuse équipe *ETE*, dont les visioconférences hebdomadaires et les projets qui lui ont survécu entretiennent encore aujourd'hui le souvenir vif des réunions de travail quotidiennes, uniques enclaves réfractaires à la dissolution temporelle du printemps 2020. Collègues géniaux de passage ou de longue date, et pour longtemps, je mesure avec émotion la chance que j'ai eu de travailler à leurs côtés au cours des trois années qui viennent déjà de s'écouler. Il s'agit de Samuel Alizon, Thomas Bénéteau, Corentin Boennec, Marc Choisy, Gonché Danesh, Ramsès Djidjou-Demasse, Morgana D'Ottavi, Baptiste Elie, Tsukushi Kamiya, Alexander Massey, Yannis Michalakis, Valérie Noël, Claudia Restrepo-Ortiz, Bastien Reyné, Quentin Richard, Christian Selinger, Olivier Supplisson et Nicolas Tessandier.



Les renforts apportés à l'équipe – décisifs mais fugaces – ont été permis par les quelques financeurs à nous avoir soutenu, à savoir la Région Occitanie, l'Université de Montpellier, l'ANRS|MIE et le CNRS. Je les en remercie ici grandement, de même que les institutions nous ayant appuyé logistiquement : l'Institut Français de Bioinformatique (IFB), l'Inserm et l'IRD (plate-forme i-Trop).

Toute la détermination qui nous animait aurait cependant été bien insuffisante pour aborder seuls l'ensemble des questions naturellement interdisciplinaires qui se sont succédées au cours de la pandémie. Je remercie ainsi les inestimables collaborations qui ont illuminé (et resteront je l'espère des phares pour certaines) ces nombreux mois de recherche et en particulier Sonia Burrel (alors Sorbonne U. & AP-HP), Vincent Foulongne & Édouard Tuillon (U. Montpellier & CHU Montpellier), Stéphanie Haim-Boukoba, Bénédicte Roquebert, & Benoît Visseaux (Laboratoire Cerba) pour la virologie, Olivier Thomine (Epidemap) & Christophe Blanchet (IFB) pour l'informatique avancée, David Morquin (CHU Montpellier) pour l'inférentielle et Jean-Yves Lefrant (U. Montpellier & CHU Nîmes) pour les soins critiques. Et puis, il y avait le providentiel 'groupe du mardi', aux origines de l'Action Coordonnée Modélisation de l'ANRS|MIE, dont je salue les membres, tout en pensant fort à Elisabeta Vergu. Cet élan est devenu une importante source de motivation.

J'exprime par ailleurs ma reconnaissance envers les collègues qui m'ont accordé leur confiance pour la suite du chemin : Nicolas Nagot, Philippe Vande Perre et Jean-Pierre Molès pour l'UMR PCEI, Éric Delaporte, Aurélie Binot et Charlotte Boullé (à qui je dois aussi la relecture de ce document) pour l'Institut ExposUM.

Merci enfin et tous les jours à celles et ceux dont le soutien absolument vital n'a pas été affaibli par cette période si particulière : à mes amis – aux premiers rangs desquels Marie-Églantine Lescasse et Théodore Soulier –, et à ma famille, qui préférerait sans aucun doute que je sois moins distrait par chaque nouveau variant décrit.



# Quantitative analyses in times of health crisis

## models for the study and the control of SARS-CoV-2 in France

Mircea T. SOFONEA, 2023

### Abstract

In spring 2020, the initial absence of immunity as well as the airborne and symptomless transmission of the SARS-CoV-2 infection led to a massive flow of COVID-19 patients in critical care units worldwide. This lightning health crisis called for rapid and accurate quantitative assessments to best inform health responses. Infectious disease modelling was however facing multiple uncertainties and a growing number of processes at play (from non-pharmaceutical interventions to vaccines and variants), varying in time and space. Hitherto less familiar with real (short) time, the field found itself having to choose between precision and robustness, objectivity and utility.

In the present summary of my work on the SARS-CoV-2 epidemic in metropolitan France, I focus on the achievements of a parsimonious modelling approach developed in March 2020 that combined the computational and tractability benefits of deterministic systems and the short-term accuracy of non-Markovian dynamics (i.e. kinetics with memory). The model was able to infer key epidemiological parameters with limited hospital data. It was then improved to continuously keep up to date with care improvement, public health policy changes and vaccine rollout. Regularly fitted with community virological data to capture viral evolution early on, it became a robust framework used to nowcast the epidemic, anticipate critical care overload at both national and local scales, and investigate counterfactual scenarios. In the light of the benefits and limitations, I discuss more broadly the place of real-time modelling during emerging infectious disease outbreak, in terms of supporting decision-making and interfacing health and research disciplines, as well as the rest of society.

Now in the post-acute pandemic era, public investment and adhesion are declining while waning immunity, immune-escape-driven evolution and the fully resumed circulation of the respiratory syncytial and influenza viruses are enabling future strain on the healthcare system. Meanwhile, the modelling frameworks, that were helpful in the first two years of the pandemic, have become unsuitable for tracking and anticipating the multiple viral dynamics (SARS-CoV-2 variants and co-circulating viruses) in a context of highly heterogeneous susceptibility. In this respect, I align my short-term research prospects with the necessary foundation of a new community-focused, joint surveillance framework of respiratory viruses, by means of updated and tailored data and quantitative analyses, primarily focusing on the polymorphic epidemiology and spatial heterogeneity challenges.

The COVID-19 pandemic has also emphasised the urgency of preparing for future health crises caused by yet unknown pathogens, generically referred to as Disease/Pathogen X. However, an actionable definition of Pathogen X and methods to evaluate the sanitary threat of an emerging pathogen in its first outbreak stage are still lacking. My long-term research perspectives will thus aim at delineating the necessary and sufficient properties shared by future pathogens likely to threaten our health system. Informing both mechanistic and phenomenological models with crowdsourced expert judgment and cross-temporal meta-analysis of past outbreak key features, this project has the ambition to develop a Pathogen X early assessment tool for long-term pandemic preparedness.

**Keywords :** COVID-19 ; hospital anticipation ; epidemiology of respiratory viruses ; viral evolution ; non-Markovian dynamics ; Pathogen X.