Epidemiological and clinical insights from SARS-CoV-2 RT-PCR crossing threshold values, France, January to November 2020

Samuel Alizon, Christian Selinger, Mircea T Sofonea, Stéphanie Haim-Boukobza, Jean-Marc Giannoli, Laetitia Ninove, Sylvie Pillet, V. Thibault, Alexis de Rougemont, Camille Tumiotto, et al.

Epidemiological and clinical insights from SARS-CoV-2 RT-PCR crossing threshold values

in press in Eurosurveillance

Samuel Alizon[1,*], Christian Selinger[1,+], Mircea T. Sofonea[1,+], Stéphanie Haim-Boukobza[2], Jean-Marc Giannoli[3], Laetitia Ninove[4], Sylvie Pillet[5], Thibault Vincent[6], Alexis de Rougemont[7], Camille Tumiotto[8], Morgane Solis[9], Robin Stephan[10], Céline Bressollette-Bodin[11], Maud Salmona[12], Anne-Sophie L'Honneur[13], Sylvie Behillil[14], Caroline Lefeuvre[15], Julia Dina[16], Sébastien Hantz[17], Cédric Hartard[18], David Veyer[19], Héloïse M Delagrèverie[20], Slim Fourati[21], Benoît Visseaux[22], Cécile Henquell[23], Bruno Lina[24], Vincent Foulongne[25], Sonia Burrel[26], SFM COVID-19 study group[27]

1. MIVEGEC, CNRS, IRD, Université de Montpellier, France
2. Laboratoire CERBA, Saint-Ouen-L'Aumône, France
3. BIOGROUP, Scientific Direction, France
4. Unité des Virus Émergents (UVE: Aix-Marseille Univ-IRD 190-Inserm 1207-IHU Méditerranée Infection), Marseille, France
5. Laboratoire des agents infectieux et d'hygiène, CHU de Saint-Etienne, France
6. Laboratoire de Virologie, CHU Rennes, France
7. Laboratory of Virology, University Hospital of Dijon, F-21000, Dijon, France
8. Univ. Bordeaux, CNRS-UMR 5234, CHU Bordeaux, Virology Department, F-33000, Bordeaux, France
9. CHU de Strasbourg, Laboratoire de Virologie, Strasbourg, France, Université de Strasbourg, INSERM, IRM UMR_S 1109, Strasbourg, France
10 Laboratoire de Microbiologie, CHU Nîmes, France
11. Nantes Université, CHU Nantes, Inserm, Centre de Recherche en Transplantation et Immunologie, UMR 1064, ITUN ; CHU Nantes, Service de Virologie; CHU Nantes, Centre d'Investigation Clinique Unité d'Investigation Clinique 1413, INSERM ; F-44000 Nantes, France
12. Laboratoire de Virologie, Hôpital Saint Louis, APHP, INSERM U976, équipe INSIGHT, Université de Paris, France
13. Assistance Publique-Hôpitaux de Paris, Hôpital Cochin, Service de Virologie, France
14. National Reference Center for Respiratory Viruses, Molecular Genetics of RNA Viruses, UMR 3569 CNRS, University of Paris, Institut Pasteur, Paris, France.
15. Département des Agents Infectieux, Laboratoire de Virologie, CHU d'Angers, Angers, France et Laboratoire HIFIH, UPRES EA 3859, Université d'Angers, Angers, France
16. Normandie Univ, UNICAEN, UNIROUEN, GRAM 2.0, 14000 Caen, France, Caen University Hospital, Department of Virology, F 14000 Caen, France
17. CHU Limoges, Laboratoire de Bactériologie-Virologie-Hygiène, F-87000 Limoges, France et Univ Limoges, RESINFIT, U 1092, F-87000 Limoges, France
18. Laboratoire de Virologie, CHRU de Nancy Brabois, Vandoeuvre-lès-Nancy, France ; Université de

Lorraine, CNRS, LCPME, F-54000, Nancy, France
19. Laboratoire de Virologie, Service de Microbiologie, hôpital européen Georges Pompidou, Assistance Publique-Hôpitaux de Paris et Unité de Génomique Fonctionnelle des Tumeurs Solides, Centre de Recherche des Cordeliers, INSERM, Université Paris, France.
20. AP-HP, Hôpital Avicenne, Laboratoire de microbiologie clinique, Bobigny, France
21. Henri Mondor Hospital, virology department, France
22. Université de Paris, Inserm, UMR 1137 IAME et Laboratoire de Virologie, Hôpital Bichat Claude Bernard, AP-HP, Paris, France
23. Service de Virologie médicale, 3IHP, CHU Clermont-Ferrand, France
24. CNR des virus des infections respiratoires (dont la Grippe), Institut des Agents Infectieux, Hopital de la Croix Rousse, HCL, Lyon, France
25. Laboratoire de Virologie, CHU de Montpellier, France
26. Sorbonne Université, INSERM U1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique (IPLESP), AP-HP, Hôpital Pitié-Salpêtrière, Service de Virologie, Paris, France
27. The member of the group are listed at the end of the article.

+ equal contribution
* author for correspondence: samuel.alizon@cnrs.fr

Abstract

The SARS-CoV-2 pandemic has led to an unprecedented daily use of molecular RT-PCR tests. These tests are interpreted qualitatively for diagnosis, and the relevance of the test result intensity, i.e. the number of amplification cycles (Ct), is debated because of strong potential biases. We analyze a national database of tests performed on more than 2 million individuals between January and November 2020. Although we find Ct values to vary depending on the testing laboratory or the assay used, we detect strong significant trends with patient age, number of days after symptoms onset, or the state of the epidemic (the temporal reproduction number) at the time of the test. These results suggest that Ct values can be used to improve short-term predictions for epidemic surveillance.

Keywords: COVID-19;SARS-CoV-2;RT-PCR;virus load;epidemiology;statistical modelling

Conflict of interest: 'None declared'.

Funding statement: This study did not receive any specific funding.

1 Introduction

Molecular testing is a key component of screening policies to control the spread of infectious diseases and the SARS-CoV-2 pandemic has led to an unprecedented testing rate using reverse transcription polymerase chain reaction (RT-PCR) assays. [1]. In clinical and public health practices, RT-PCR results are qualitative for viral respiratory disease diagnostics, with reports such as 'positive', 'negative', 'uninterpretable', and, sometimes, 'weakly positive'. These are based on the cycles threshold, also referred to as crossing point or crossing threshold (here denoted C_t), which corresponds to the number of PCR amplification cycles required for the fluorescent signal to rise above a positive threshold. In theory, the more abundant the genetic target in the sample, the fewer the amplification cycles required to detect it. This is why numerous studies on SARS-CoV-2 rely on C_t values to assess transmissibility [2] or study infection kinetics [3]. However, many practical and biological limitations make C_t values a poor reflector of virus load [6]. Here, we present a cross-sectional analysis of SARS-CoV-2 RT-PCR tests performed on 2,220,212 individuals in France between January 21, 2020, and November 30, 2020 (Figure S2).

Few studies analyze C_t values at a population level. One explanation for this matter of affairs is that these are known to suffer from several, potentially strong, biases. First, sample type and sampling quality directly affect the amount of genetic material available. Second, the RT-PCR assay matters. Even the quality of the reagents used may have a significant effect on the number of amplification cycles required to achieve the same level of fluorescence for the same amount of target genetic material. Combining data from different laboratories helps to control for these sources of variation in statistical analyses. Furthermore, the larger the dataset, the more we can detect small statistical trends even after having controlled for non-informative variables.

2 Methods

In our analysis, we studied tests from individuals aged between 1 to 89 years old performed in France in 2020. There currently is no national database so datasets from the 21 partner laboratories were combined for this study. The context in which these tests were performed varied temporally. Until at least April 2020, the testing capacity was limited in the country and the majority of tests were performed on symptomatic individuals, especially in hospital settings. After May 2020, testing was more accessible and data then includes screening tests performed in the general population. This change in testing context coincided with a shift in terms of screening facilities with the majority of the tests being performed in hospital virology departments until April 2020 and in private laboratories after that. Conversely, we do not expect this testing motivation to vary across French regions. We did not take into account tests for which key variables such as patient age, patient sex, laboratory geographical department, qualitative result, or RT-PCR assay used were unknown. Note that one test could provide more than one C_t value if containing probes targeting multiple viral genes. According to the national guidelines [4], it is recommended to focus on the most sensitive target to categorize levels of viral excretion. After removing the 388 biologically unrealistic C_t values that were lower than 10 or larger than 45, the 95% confidence interval (95CI) of the remaining values was [16.89;35.56] (Table **S1**). The median and upper bound of the 95CI were unaffected by the removal of these values and the lower bound increased from 16.87 to 16.89. Overall, we were left with 793,479 tests from the same number of individuals. These originated from both negative and positive tests but, since they all have C_t values, the former are less represented.

We used a linear regression model to explore how C_t values can be explained by the following variables: patient age and sex, the number of days since the onset of the symptoms (if known), the clinical sampling site (if known), the sampling facility (if known), the RT-PCR assay used, the target gene, the test's qualitative result, the sampling date, the temporal reproduction number of the epidemic at the sampling date (denoted R_t and estimated on national hospital admission data using the EpiEstim method [5]), and a control variable. The latter corresponds to the last digit of the patient anonymity number and is expected to be independent of the C_t value. Therefore, the lowest p-value associated with the control variable, which we expect to be pure noise, can be used to set the significance threshold for the other variables. We also included in the model an interaction term between sampling date and R_t. For this analysis, we excluded C_t values from internal controls. Univariate analyses are extremely sensitive to heterogeneity in the dataset. For instance, the age distribution from patients sampled in aged care homes is different than that from city screening facilities, and analyzing the 'sampling facility' factor alone could yield misleading results. This is why the analysis used here is multivariate and considers all the factors listed above simultaneously. In particular, allows us to control for variations in the way the data was collected, e.g. the intensity or the context of the sampling.

To control for the consistency of the results for some of the factors, especially those related to the infection (e.g. the number of days since symptoms onset), we also performed the analyses only on the tests that were reported as "positive" or "weakly positive". These are shown in Supplementary Results.

3 Results

The adjusted R-square of the linear model was 38.8%, meaning that the factors we chose explained one-third of the variance in C_t values. The model residuals were normally distributed (Figure S3A). Care should be taken in the analysis given that the data is unbalanced, which is why we performed an analysis of variance (ANOVA) with type II sums of squares. All the factors except the temporal reproduction number were significantly associated with C_t values using a classical 5% p-value criterion. Even for the control variable, the p-value was 0.013 and patients with final digits 1 and 3 had C_t values slightly lower (-0.19 and -0.17 cycles) than patients with a 0 final digit. Therefore, we set our significance thresholds to 5% of that of the control variables, *i.e.* 6.5×10^{-4}, to analyze the main effects (Table 1). Detailed outputs of the linear model are shown in Table S2.

The intercept of the linear model indicates the average C_t value for a positive test performed with the reference assay, and all the other factors being set to their reference value. Its magnitude (19.1 cycles) is in line with clinical practice. The importance of the noise in the dataset is illustrated by the strong effect of the testing laboratory, as well as the RT-PCR assay used (Supplementary Figure 2).

Despite this high level of noise, we detected a strong effect of the qualitative result (Figure 1A), with C_t differences that were even larger than that from the laboratory effect. We also found a slightly significant difference of -1.81 cycles between the most common type of samples (nasopharyngeal) and that performed on other clinical sampling sites (mostly lower respiratory tracts, but also faeces or saliva). This is likely because the latter tests were performed in patients with more severe symptoms.

The effects associated with the number of days since symptoms onset was particularly strong. For the 8.5% of the participants for whom the number of days between symptoms onset and testing dates was known, we found that the C_t gradually increases over the reported range with a maximum difference of 5.73 cycles (Figure 1B). The effect was similar when removing the tests clinically considered to be "negative" from the analysis (Supplementary Table S3).

The effect of sex had the same order of magnitude as that of the control variable and could, therefore, be treated as non-significant. Conversely, the age factor had a strong effect with a decrease of 0.541 cycles per year (Figure 1C).

The target gene of the RT-PCR assay used also yielded a slightly significant effect. The C_t values obtained when using a probe targeting the ORF1 and S regions of the virus genome were significantly higher than when using the N gene, which was the genomic region of reference in the model (Figure 1D). This effect is consistent with the life-cycle of the virus. As stressed by [6], since coronaviruses are (+)ssRNA viruses, they use the same RNA matrix for replication and transcription, both being amplified by diagnostic assays. Furthermore, *Coronaviridae* transcripts can produce subgenomic mRNAs that lack part of the genome [7]. As a consequence, and as shown in cell cultures [8], genes on the 5' end of the genome are under-represented. This is consistent with our result where assays targeting the gene on the 3' end (the N gene) tend to have lower C_t than assays targeting genes on the 5' end (the ORF1 and S genes). Note that an alternative explanation could be that some probes target more conserved areas of the SARS-CoV-2 than others [9].

Finally, we found that C_t values decreased with time (-0.797 cycle per day), but this effect was non-linear (Figure 1E). This could be due to the strong variation in testing efforts in France (Figure S2A), but also to variations in the epidemic trend. Indeed, although the R_t (inferred from hospitalization data using the EpiEstim method [5]) was not found to be significant, the interaction between the sampling date and R_t was nearly significant (Figure 1F), suggesting that a temporal analysis could yield additional insights.

The existence of a correlation between the C_t values of the tests performed in a population and R_t is consistent with population dynamics theory, which predicts that in an expanding population of infected individuals, the 'age' of the infections, *i.e.* the number of days post-infection, is skewed towards lower values [10]. Since C_t values have been reported to increase over the course of an infection [3], which we confirm with this analysis (Figure 1D), it has been suggested that these values could be used as an early signal to predict R_t [11].

To investigate this question, we focused on screening data collected in the general population from individuals aged from 5 to 79. We estimated the median and skewness values of the daily distribution of the C_t values. To correct for potential confounding factors, these were adjusted using a linear model (see the Supplementary Methods). We analyzed the temporal correlation between the time series with a 7-day rolling average of this median, skew, and R_t (Figure 2). For the median C_t value, we found a significant correlation with R_t that was maximized for a 6 to 7 days delay (Figure **S4**). This is consistent with R_t being calculated using data from ICU-admissions, which occur with a median of 14 days after infection [12, 13], and RT-PCR screening data being obtained earlier in the infection.

To further assess the usefulness of C_t data, we used ARIMA models to predict R_t dynamics over 7 days. We compared models without any exogenous data, to models that also included exogenous time series (either median or skewness of estimated C_t values distribution, or the fraction of positive tests [1]). As expected, the prediction error made using only endogenous data (R_t) was low in periods where R_t variations were limited. Furthermore, we found that adding exogenous data improved the prediction, especially when strong shifts in R_t were occurring (Figure 2). C_t values (green and cyan dots) tended to provide a better reduction in the error of the prediction than the ratio of positive tests (purple).

4 Discussion

This analysis of a large national database of RT-PCR tests performed in the context of a major epidemic confirms that population-level C_t values are noisy since even a linear model that features 91 degrees of freedom does not explain the majority of the variance. However, owing to the law of large numbers, we detect several effects that are in line with biological observations and with virological properties. For instance, our finding that C_t values decrease as a function of the number of days after symptoms onset is consistent with longitudinal follow-ups [3]. Another study also reported lower C_t values when the test was performed in symptomatic individuals [17]. This same found that men had slightly lower C_t values than women, which was not significant in our analysis. Similarly, the difference we detect between the virus gene targeted by the RT-PCR assay used can be interpreted in the light of known differences in mRNA copy numbers between genes depending on their distance from the 3' end [8]. Concerning the link between age and C_t values, although there are some mechanistic interpretations as to why virus load would increase with age [14], the evidence was mixed, with some studies reporting a decreasing trend [15] and others not [16, 17]. Here, using a multivariate approach on a large dataset allows us to unravel a strong and significant decrease of C_t values with age.

A limitation of our study is that although our dataset stands out by its size and its level of details, it is restricted to a single country where testing effort varies, both on a temporal and on a spatial scale (Figure S2). Performing similar analyses in other European countries can therefore be particularly informative. In the study, we also chose to analyse all the tests performed that had Ct value. This is debatable since high Ct values can be due to noise and this is the point of implementing cutoffs. However, Ct values are also known to increase of the course of an infection [3] and these high values could also correspond to patients detected in a late stage, which is expected to be more frequent in a declining epidemic [11]. To control for this potential bias, we also performed the analysis on a dataset without the tests with a “negative” result. Finally, this analysis was conducted at the end of the year 2020 but since then, as in most countries, the emergence of variants has dramatically altered epidemiological dynamics in France [19,20] and early reports suggest that the Ct value measured could depend on the variant causing the infection [21,22,23,24,25]. Vaccination has also greatly changed the picture as indicated by Ct estimations in vaccine breakthrough infections [24,25] and should be included as a host factor in future analyses.

As pointed out elsewhere, care should be taken when interpreting C_t values because of technical issues (different assays may yield higher or lower values) and of biological issues (coronaviruses produce subgenomic RNAs) [6]. However, in this analysis, we do not attempt to link C_t values to viral loads but rather analyse raw values at a population level. A promising output of this analysis is the possibility to use C_t values as an early signal to detect changes in epidemic behaviour, *e.g.* R_t values. Indeed, our most robust descriptors of the epidemics originate from hospital-admission data, but these have a significant delay with the status of the epidemic since patients are hospitalized 2 weeks after infection [13, 12]. The ratio of positive tests performed in the population of interest can, in theory, provide earlier insights but it suffers from strong sampling biases. We show that accounting for population-based C_t values variations can improve R_t predictions on a 7-days period. This result is consistent with that obtained recently using data from nasopharyngeal specimens collected from staff and residents in four Massachusetts long-term care facilities [11]. Note that, contrarily to us, this earlier study does not factor in individual data regarding patient age or symptomatic status and it does not perform a cross-validation analysis that would control for temporal autocorrelation issues.

Our results show that analysing a large dataset of C_t values from screening tests allows filtering out the important amount of noise in these values. Their inclusion in routine surveillance calls for an adaptation to the current state of the epidemics, especially the evolution of variants and the increase in vaccination coverage, but also the integration with other types of data such as mobility data [26,27,28]. They also call for better integration of C_t values in national and European public health policies to monitor epidemics caused by SARS-CoV-2 but also other human viruses, especially since these data raise less ethical concerns than other sources of data such as mobility data.

References

[1] Max Roser EOO Hannah Ritchie, Hasell J. Coronavirus pandemic (COVID-19). Our World in Data. 2020;.

[2] 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.

[3] Néant N, Lingas G, Hingrat QL, Ghosn J, Engelmann I, Lepiller Q, et al. Modeling SARS-CoV-2 viral kinetics and association with mortality in hospitalized patients from the French COVID cohort. PNAS. 2021;118(8):e2017962118.

[4] French Microbiology Society (SFM). Avis du 25 septembre 2020 de la Société Française de Microbiologie (SFM) relatif à l'interprétation de la valeur de Ct (estimation de la charge virale) obtenue en cas de RT-PCR SARS-CoV-2 positive sur les prélèvements cliniques réalisés à des fins diagnostiques ou de dépistage; 2021. Available from: https://www.sfm-microbiologie.org/wp-content/uploads/2021/01/Avis-SFM-valeur-Ct-excre%CC%81tion-virale-_-Version-def-14012021_V4.pdf

[5] Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. Am J Epidemiol. 2013 Nov;178(9):1505–1512.

[6] Michalakis Y, Sofonea MT, Alizon S, Bravo IG. SARS-CoV-2 viral RNA levels are not "viral load". Trends Microbiol. 2021; in press, doi:10.1016/j.tim.2021.08.008

[7] Ogando NS, Dalebout TJ, Zevenhoven-Dobbe JC, Limpens RWAL, van der Meer Y, Caly L, et al. SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. Journal of General Virology. 2020;101(9):925–940.

[8] Finkel Y, Mizrahi O, Nachshon A, Weingarten-Gabbay S, Morgenstern D, Yalahom-Ronen Y, et al. The coding capacity of SARS-CoV-2. Nature. 2021;589(7840):125–130.

[9] Dearlove B, Lewitus E, Bai H, Li Y, Reeves DB, Joyce MG, et al. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. PNAS. 2020;117(38):23652-23662.

[10] Caswell H. Matrix population models: construction, analysis and interpretation. MA 01375 USA: Sinauer Associates, Inc.; 1989.

[11] Hay JA, Kennedy-Shaffer L, Kanjilal S, Lipsitch M, Mina MJ. Estimating epidemiologic dynamics from single cross-sectional viral load distributions. Science. 2021, DOI: 10.1126/science.abh0635

[12] 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. Epidemics. 2021;35:100459.

[13] Salje H, Kiem CT, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, et al. Estimating the burden of SARS-CoV-2 in France. Science. 2020;369:208-211.

[14] Yang Y, Zhao Y, Zhanga F, Zhang L, Li L. COVID-19 in Elderly Adults: Clinical Features, Molecular Mechanisms, and Proposed Strategies. Aging and disease. 2020;11(6):1481–1495.

[15] Euser S, Aronson S, Manders I, Lelyveld Sv, Herpers B, Sinnige J, et al. SARS-CoV-2 viral load distribution reveals that viral loads increase with age: a retrospective cross-sectional cohort study. medRxiv. 2021;p. 2021.01.15.21249691.

[16] Jones TC, Mühlemann B, Veith T, Biele G, Zuchowski M, Hofmann J, et al. An analysis of SARS-CoV-2 viral load by patient age. medRxiv. 2020;p. 2020.06.08.20125484.

[17] Walker AS, Pritchard E, House T, Robotham JV, Birrell PJ, Bell I, et al. Ct threshold values, a proxy for viral load in community SARS-CoV-2 cases, demonstrate wide variation across populations and over time. eLife. 10:e64683. DOI: 10.7554/eLife.64683

[18] Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics. 2019;29:100356.

[19] Haim-Boukobza S, Roquebert B, Trombert-Paolantoni S, Lecorche E, Verdurme L, Foulongne V, et al. Detection of Rapid SARS-CoV-2 Variant Spread, France, January 26--February 16, 2021. Emerging Infectious Diseases. 2021;27(5):1496–9.

[20] 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 15;26(28):2100573.

[21] 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;593:270–4.

[22] Faria NR, Mellan TA, Whittaker C, Claro IM, Candido D da S, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. Science. 2021;372(6544):815–21.

[23] Roquebert B, Haim-Boukobza 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. medRxiv. 2021;2021.03.19.21253971.

[24] Brown CM. Outbreak of SARS-CoV-2 Infections, Including COVID-19 Vaccine Breakthrough Infections, Associated with Large Public Gatherings — Barnstable County, Massachusetts, July 2021. MMWR Morb Mortal Wkly Rep. 2021;70. https://www.cdc.gov/mmwr/volumes/70/wr/mm7031e2.htm

[25] Blanquart F, Abad C, Ambroise J, Bernard M, Cosentino G, Giannoli J-M, et al. Characterisation of vaccine breakthrough infections of SARS-CoV-2 Delta and Alpha variants and within-host viral load dynamics in the community, France, June to July 2021. Eurosurveillance. 2021;26(37):2100824.

[26] Kraemer MUG, Hill V, Ruis C, Dellicour S, Bajaj S, McCrone JT, et al. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. Science. 2021 Aug 20;373(6557):889–95.

[27] Pullano G, Di Domenico L, Sabbatini CE, Valdano E, Turbelin C, Debin M, et al. Underdetection of cases of COVID-19 in France threatens epidemic control. Nature. 2021 Feb;590(7844):134–9.

[28] Selinger C, Choisy M, Alizon S. Predicting COVID-19 incidence in French hospitals using human contact network analytics. International Journal of Infectious Diseases. 2021; doi:10.1016/j.ijid.2021.08.029

Table 1: Main factors affecting Ct values in the multivariate linear model. We only list factors with significant effects with a 10^{-3} p-value criterion. Coefficients reflect differences in C_t. For qualitative factors, the reference value (denoted "ref.") is shown. CI: confidence interval.

Factor	Value	Coefficient	2.5% CI	97.5% CI
(intercept)		19.1	12.9	25.4
assay	PerkinElmer (ref.)	—	—	—
	Genefinder	12.1	10.3	13.9
laboratory	LAB_1 (ref.)	—	—	—
	LAB_122	5.42	3.79	7.05
	LAB_96	-4.8	-6.71	-2.90
result	positive (ref.)	—	—	—
	weakly positive	11.3	11.1	11.5
	negative	16.9	16.6	17.2
days post-symptoms onset	less than 4 (ref.)	—	—	—
	4 to 7	2.76	2.66	2.86
	8 to 14	4.90	4.73	5.08
	more than 14	5.73	5.43	6.03
sample	naso-pharyngeal (ref.)	—	—	—
	other	-1.81	-2.49	-1.14
age	(per year)	-0.541	-0.585	-0.497
target gene	N (ref.)	—	—	—
	ORF1	1.03	0.949	1.12
	S	1.19	0.948	1.43
date	(per day)	-0.797	-0.903	-0.691

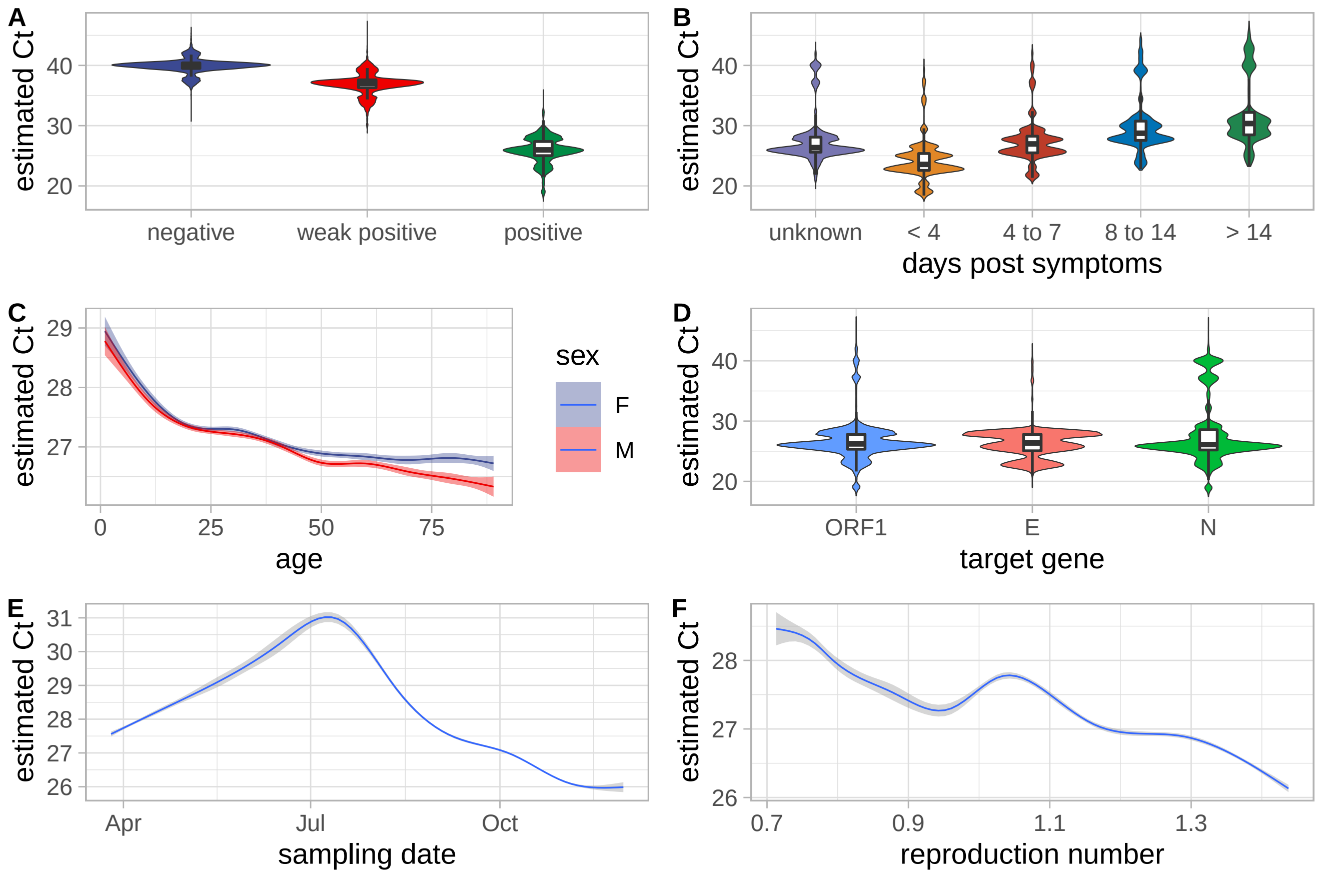

Figure 1: **Correlations between key factors and observed *Ct* variations.** A) Qualitative result of the test, B) number of days between symptoms onset and testing, C) participant age and sex, D) genomic area targeted by the test, E) sampling date, and F) temporal reproduction number (R_t) at the time of the test. For panels A, B, and D, the violin plots indicate the distributions and the box plots show the 0.025, 0.25, 0.5, 0.75, and 0.975 quantiles. Panels C, E, and F are obtained with a 'loess' smoothing model and the gray area shows the 95% confidence interval from the underlying model. The C_t values shown are not the raw values but that estimated using a multiparametric linear model to correct for biases (see Supplementary Methods for details).

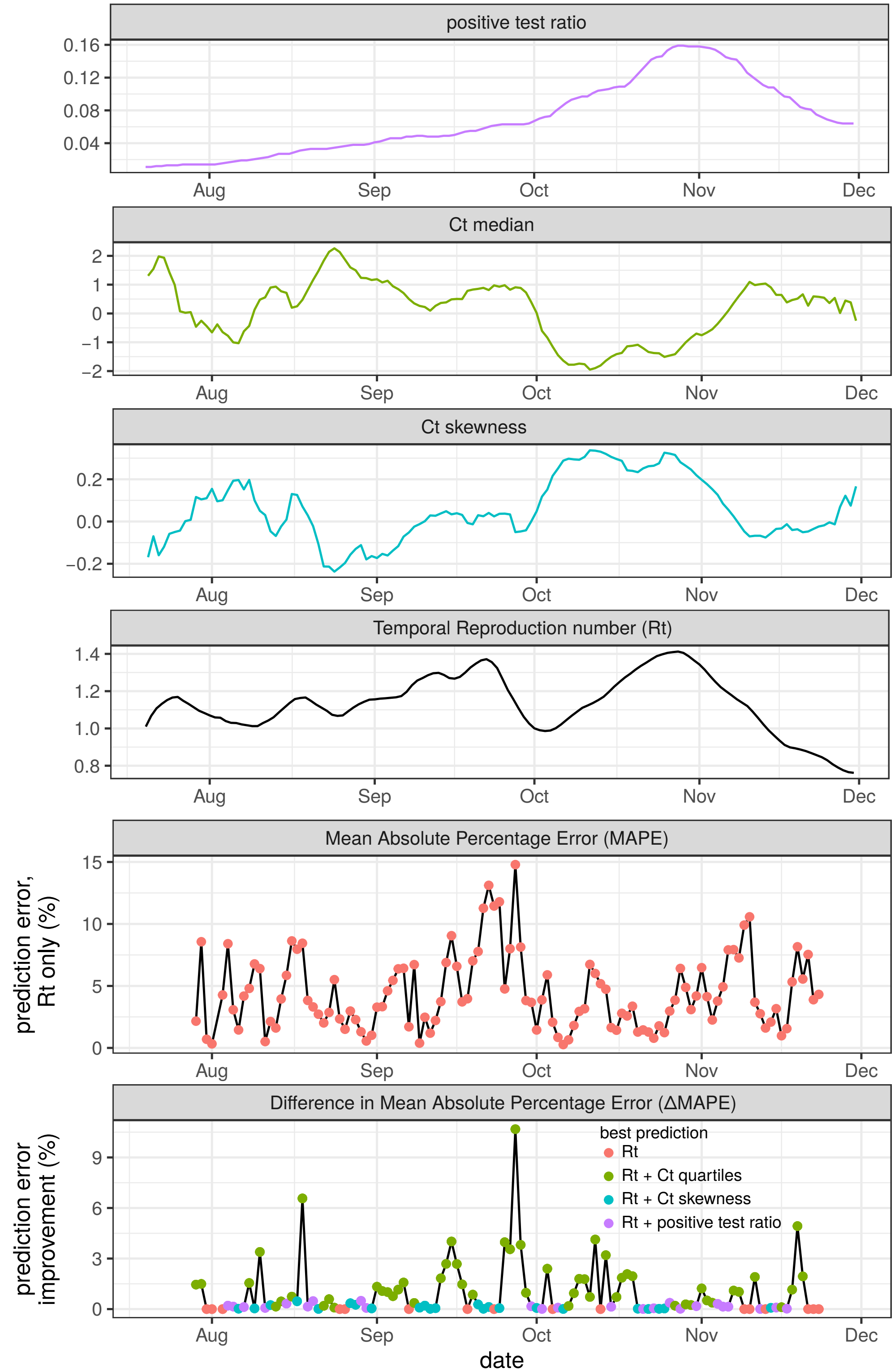

Figure 2: **Predicting temporal reproduction number (R_t) from time series.** The top four panels show the the 7-days rolling averages of the time series of the ratio of positive tests (in purple), the median (in green) and skewness (in cyan) of the daily C_t residual distribution, and R_t (in black). The bottom panels show the error made by a prediction using only R_t data (red dots) and the potential improvement made by including exogeneous data.