

# Modelling The Positive Testing Rate of Covid-19 in South Africa Using a Semi-Parametric Smoother for Binomial Data

**Olajumoke Evangelina Owokotomo**

Hasselt University

**Samuel Manda**

South African Medical Research Council

**Adetayo Kasim**

Durham University

**Jürgen Claesen**

Hasselt University

**Tarylee Reddy** (✉ [tarylee.reddy@mrc.ac.za](mailto:tarylee.reddy@mrc.ac.za))

South African Medical Research Council

**Ziv Shkedy**

Hasselt University

---

## Research Article

**Keywords:** COVID-19, South Africa, Number of tests, Infection rate, Positive testing rate, Smoothing Binary data

**Posted Date:** March 10th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-228727/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

## RESEARCH

# Modelling the positive testing rate of COVID-19 in South Africa Using A Semi-Parametric Smoother for Binomial Data

Olajumoke Evangelina Owokotomo<sup>1</sup>, Samuel Manda<sup>2,4,5\*</sup>, Adetayo Kasim<sup>3\*</sup>, Jürgen Claesen<sup>1</sup>, Tarylee Reddy<sup>1,2,4\*</sup> and Ziv Shkedy<sup>1</sup>

\*Correspondence:

[samuel.manda@mrc.ac.za](mailto:samuel.manda@mrc.ac.za);

[a.s.kasim@durham.ac.uk](mailto:a.s.kasim@durham.ac.uk);

[tarylee.Reddy@mrc.ac.za](mailto:tarylee.Reddy@mrc.ac.za)

<sup>2</sup>Biostatistics Research Unit, South African Medical Research Council, Capetown, South Africa.

<sup>4</sup>School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal,

KwaZulu-Natal, South Africa.

<sup>5</sup>Department of Statistics, University of Pretoria, Pretoria, South Africa.

<sup>3</sup>Department of Anthropology, Durham Research Methods Centre, Durham University, Durham, United Kingdom.

<sup>1</sup>Data Science Institute, Center for Statistics, I-BioStat, Hasselt University, Hasselt, Belgium.

Full list of author information is available at the end of the article

## Abstract

**Background:** The current outbreak of COVID-19 is a major pandemic that has adversely affected the world economies, societies and also an increase in health burdens within a short time. South Africa has the highest number of confirmed COVID-19 cases in Africa and is the fourteenth most affected country in the world. Understanding the country's COVID-19 infection rate will help in preventing the spread of the disease.

**Method:** We propose to estimate both the the positive testing rate and the rate at which the positive testing rate changes over time using a flexible semi-parametric model. We used publicly available data collected from March 5th to September 2nd 2020.

**Results:** We found that the positive testing rate was declining from early March when the disease was first observed until early May. In the month of July 2020, the infection reached its peak then its started to decrease again.

**Conclusion:** The observed increase in the positive testing rates between May and July 2020 could imply that the testing algorithm was effective; however its effectiveness declined after the end of July 2020, which coincided with the end of winter season in the country.

**Keywords:** COVID-19; South Africa; Number of tests; Infection rate; Positive testing rate; Smoothing Binary data

## Background

Coronaviruses are a large family of viruses which may cause respiratory infections ranging from the common cold to more severe diseases such as MERS and SARS. The ongoing outbreak of the novel coronavirus SARS-CoV-2 was first detected on 31st December 2019 in Wuhan, China [1]. The virus has rapidly spread with a total of 37,423,660 confirmed cases and 1,074,817 deaths as of 12th October 2020 [1].

The South African government declared a national state of disaster on March 15th 2020 and commenced a state of lockdown from March 26th 2020 in an effort to reduce COVID-19 transmission in the country [2]. During this period all international and inter-provincial borders were closed, as well as the education sector and several economic sectors in the country. As of June 2020, the country adopted a COVID-19 risk-adjusted strategy with a phased re-opening of selected economic sectors and schools.

Modeling the number of COVID-19 cases and in particular producing a reliable

short and long term predictions of the number of COVID-19 cases are critical tools for policy makers to design interventions in order to control the disease's spread. Recently, [3] applied a robust model based approach, that does not require making assumptions about the transmission process to model the number of COVID-19 cases and they were able to provide accurate short term prediction for 5-10 days using the South Africa COVID-19 data. These non-linear epidemiological models have previously been applied to model other disease outbreaks such as Ebola [4], Dengue [5], Zika virus [6] and, more recently, the COVID-19 pandemic [7]; [8]; [9]. Specifically, [7] fitted the generalized logistic model, Richards's model and a sub-epidemic model to the cumulative COVID-19 cases in the Hubei province of China and produced a short-term forecast of 5, 10 and 15 days ahead. In a recent analysis by [8], a similar approach was used to estimate the key epidemic parameters for all 11 provinces in China as well as 9 selected countries. All the models discussed above made use of the daily or cumulative number of cases to fit the models and to estimate the parameters of interest. In the context of COVID-19, this introduces a difficulty as seen in Figure 1, since in South Africa (and many other countries) the number of tests and number of cases are correlated [3].

Positive testing rate, i.e., the probability of positive COVID-19 cases among the total number of COVID-19 tests, has been seen as an important metrics in understanding the transmission of COVID-19 in the literature [10]. Due to the correlation (dependence) between the number of COVID-19 cases and the number of COVID-19 tests conducted, no country knows the actual total number of people infected with COVID-19 but only the infection status of those who have been tested. Therefore, in countries with a high positive rate, the number of confirmed COVID-19 cases is more likely to represent only a small proportion of the true number of cases. However, when the positive rate increases it can suggest the virus is actually spreading faster than the growth seen in confirmed cases. The positivity rate is of great importance, and is used to guide policy makers on COVID-19 interventions. This was evidenced on May 12, 2020 when the WHO advised governments that before relaxing intervention measures, rates of positivity in testing should remain at 5% or lower for at least 14 days (John Hopkins coronavirus resource center, 2020, WHO, 2020).

[11] recently explored the relationship between demographic factors and the testing positivity rate in specimens from a particular hospital in Wuhan, China. Other authors, through examination of the daily COVID-19 incidence and testing, showed that changes in testing rates could mask the epidemic's growth rate, which has public health implications [12]. While several COVID-19 dashboards have included reporting of the positive testing rate at the national and provincial levels [1].

Recently, authors have attempted to estimate the state-level COVID-19 prevalence in the United States using COVID-19 test positivity and reported COVID-19 case rates [13]. To our knowledge there have been limited efforts to directly model the COVID-19 positive testing rate and the rate of change in this area over time. To provide a more accurate perspective on the disease burden, we propose an alternative modeling approach that focuses on COVID-19 positive testing rate, i.e., the probability of positive cases per tests conducted and the rate of change in this rate over time. In this paper we model the daily number of COVID-19 cases among

the number of COVID-19 tests carried out using a semi-parametric model in which the rate of change of the positive testing rate is estimated using a smooth function of time. In particular, we apply scatter plot smoothing techniques for binomial data using generalized additive models in order to obtain an estimate of the rate of change [14]. In Section we describe the testing policy in South Africa from which the data used for the analysis presented in this paper was obtained. The modelling approach, the model formulation for the positive testing rate and the methodology to construct simultaneous confidence bands are discussed in Section . Section contains the results, and the discussions and conclusions are in Section .

## Data

### Daily number of tests and confirmed cases

The daily number of reported COVID-19 cases and tests for the period of March 7th 2020 to September 2nd 2020 is presented in Figure 2. The growth of COVID-19 infections in South Africa appears to be tri-phasic especially during the early phase when the cumulative cases were low with rapid growth until March 27th 2020. A total of 243 daily new cases were observed, followed by a sharp decline in the rate of new cases. From March 28th 2020 to April 6th 2020 the daily increase in cases was consistently below 100. From May 2020 onwards, a consistent increase of more than 1000 cases per day were observed. The peak period was between of July 9th and 19th where more than 10,000 reported cases were reported on a daily basis. As of July, a total of 3726721 tests had been conducted, corresponding to a testing rate of 22.816 per 1000 population. Throughout this period, the proportion of infections increased until mid July when it started to decrease.

### Testing policy in South Africa

A total of 3,245,087 tests for SARS-COV-2 were conducted between March 1st and August 29th 2020. These tests were performed on individuals who satisfied the case definition for PUI. The PUI definition, which was amended consistently included at least one of the following criteria: symptomatic individuals seeking testing, hospitalized individuals for whom testing was done, individuals in high-risk occupations (e.g health care workers), individuals in outbreak settings, and individuals identified through community screening and testing programmes which were implemented between April and the middle of May 2020. The number of tests performed on a weekly basis increased from March 2020 until the third week of May and proceeded to decrease over the subsequent two weeks due to a limited supply of testing kits. The average time elapsed from specimen collection to testing was under two days in both the private and public sectors from August 22th to 29th August 2020.

## Modeling COVID-19 infection rate in South Africa using Generalized Linear Mixed Effects Model for Binary Data

### Model formulation for the positive testing rate

The number of positive cases is assumed to be binomially distributed. Let  $\pi_t$  be the daily positive testing rate per test,  $Y_t$  be the daily number of COVID-19 cases and  $n_t$  be the daily number of COVID-19 tests. Our aim is to model the probability  $\pi_t$  and to produce a model-based estimate for its first derivative, i.e., the change in

the positive testing rate over time. Semi-parametric regression model for binomial data was used to provide an estimate of the positive testing rate as a function of time. The relationship can be expressed as

$$\begin{aligned} Y_t &\sim \text{Bin}(n_t, \pi_t), \quad t = 1, \dots, T, \\ \text{logit}(\pi_t) &= f(t). \end{aligned} \quad (1)$$

Here,  $f(t)$  is a smooth function of the time  $t$ . Smoothing splines are commonly used for this purpose ([14]). A general spline model of degree  $d$  with  $K$  knots can be written as follows:

$$\text{logit}(\pi_t) = \beta_0 + \beta_1 x_i + \dots + \beta_d x_i^d + \sum_{k=1}^K u_k s_k(x_i), \quad (2)$$

where  $s_k(x)$  is a set of spline basis functions.

To avoid overfitting, the spline model is typically estimated by considering penalized maximum likelihood estimation, with a penalty term of the form  $\lambda \sum_k u_k^2$ . [14] showed that the penalized regression problem can be expressed as an equivalent GLMM:

$$\text{logit}(\pi) = \mathbf{X}\beta + \mathbf{Z}\mathbf{u}, \quad (3)$$

with  $\mathbf{p}\mathbf{i} = [\pi_1, \pi_2, \dots, \pi_T]^T$ ,  $\mathbf{beta} = [\beta_0, \beta_1, \dots, \beta_d]^T$ , and  $\mathbf{u} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_K]^T$ . Note that  $\mathbf{beta}$  and  $\mathbf{u}$  are vectors of the fixed and random effects, respectively, with  $u_k \sim \mathcal{N}(0, \sigma_u^2)$  where  $\sigma_u^2$  acts as the smoothing parameter. This representation has the advantage that the degree of smoothing can be estimated from the data using standard mixed-model software (e.g., [14], chapter 4). The design matrices  $\mathbf{X}$  and  $\mathbf{Z}$  are defined as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & \dots & x_1^d \\ 1 & x_2 & \dots & x_2^d \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_T & \dots & x_T^d \end{bmatrix}$$

and

$$\mathbf{Z} = \begin{bmatrix} s_1(x_1) & s_2(x_1) & \dots & s_K(x_1) \\ s_1(x_2) & s_2(x_2) & \dots & s_K(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ s_1(x_T) & s_2(x_T) & \dots & s_K(x_T) \end{bmatrix}.$$

The estimation of the model (3) is performed by means of penalized quasi-likelihood (PQL). Initial estimates for  $\mathbf{beta}$  and  $\mathbf{u}$  are used to calculate the pseudo-data  $\mathbf{y}^*$ :

$$\mathbf{y}^* = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{W}^{-1}(\mathbf{y} - \boldsymbol{\pi}) \equiv \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}^*, \quad (4)$$

where  $\mathbf{W}$  is a diagonal matrix with variances of  $y_i$  on the diagonal. The pseudo-error  $\boldsymbol{\varepsilon}^*$  has a variance-covariance matrix  $\mathbf{R} = \mathbf{W}^{-1}\phi$ , where  $\phi$  is the dispersion parameter, equal to one for the standard binomial model family. Equation (4) resembles a LMM formulation for  $\mathbf{y}^*$ . Thus, an LMM is fitted to the pseudo-data, yielding updated estimates of  $\boldsymbol{\beta}$ ,  $\mathbf{u}$ ,  $\sigma_u^2$ , and  $\phi$ . The procedure of calculating pseudo-data and re-fitting the LMM is repeated until convergence.

#### Estimating the change in the positive testing rate

To understand the change in positive testing rates over time, we used a simple rate of change as: the rate of change in the positive testing rate over time using the derivative of  $\pi_t$  given by

$$\pi'_t = \frac{\pi(t) - \pi(t-1)}{\Delta(t)}. \quad (5)$$

It is assumed that if the number of tests is constant over time and applied to a random sample of the population,  $\pi'_t$  can give an indication to the change in the virus transmission in the population (since in this case, it gives the change in transmission probability). However, it is unlikely to assume that the number of tests will be constant nor that the tests will be applied to random sample from the population. Also in this case, the derivative can provides a good indication about the general trend of the virus' transmission for the tested population and can be used as a tool to asses the success of an implemented intervention strategy.

#### Construction of pointwise confidence band

According to [14], an approximate  $100(1-\alpha)\%$  pointwise confidence band for an estimated penalized spline in the GLMM framework,  $\hat{f}(x)$ , is given by:

$$\hat{f}(x) \pm z_{1-\alpha/2} \times \widehat{\text{st.dev}}\{\hat{f}(x) - f(x)\}, \quad (6)$$

where

$$\widehat{\text{st.dev}}\{\hat{f}(x) - f(x)\} = \sqrt{C_x \hat{Q} C_x^T}, \quad (7)$$

with  $C_x = (1 \ x \ \dots \ x^d \ s_1(x) \ \dots \ s_K(x))$  and

$$\hat{Q} = \widehat{\text{cov}} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} - \mathbf{u} \end{bmatrix} = (C^T \hat{R}^{-1} C + 1/\hat{\sigma}_u^2 D)^{-1}, \quad (8)$$

where  $C = [XZ]$  and  $D \equiv \text{diag}([0_{d+1}^T, 1_K^T])$

Pointwise confidence bands, however, need to be corrected for multiplicity. Also, they ignore serial correlation. Therefore, we make use of simultaneous confidence bands implemented in [15], which allow to make joint statements on multiple locations of the fitted curve. A  $100(1-\alpha)\%$  simultaneous confidence band for  $\hat{f}_x$  is defined as:

$$\hat{f}_x \pm c_{1-\alpha} \times \widehat{\text{st.dev}}\{\hat{f}(x) - f(x)\} \quad (9)$$

where the critical value,  $c_{1-\alpha}$ , is the  $(1-\alpha)$  quantile of the random variable

$$\sup_{x \in \mathcal{X}} \left| \frac{\widehat{f}(x) - f(x)}{\widehat{\text{st.dev}}\{\widehat{f}(x) - f(x)\}} \right| \approx \max_{1 \leq l \leq M} \left| \frac{\left( C_x \begin{bmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\mathbf{u}} - \mathbf{u} \end{bmatrix} \right)_l}{\widehat{\text{st.dev}}\{\widehat{f}(x_l) - f(x_l)\}} \right|,$$

which can be found by simulating from an approximate multivariate normal distribution.[14]:

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\mathbf{u}} - \mathbf{u} \end{bmatrix} \sim \mathcal{N} \left\{ \mathbf{0}, \widehat{\mathbf{Q}} \right\}.$$

### Application to the data

A generalized additive model was fitted to the data with the time component as the smooth term using the `gam()` function of the `mgcv` [16] library in R [17]. Figure 3, shows that the estimated positive testing rate reached its peak on July 19th, 2020, at the same time that the number of tests was at its highest level. From that time onward, both number of tests and the positive testing rate declined. This could be a result of a reduction of the virus' transmission in the population or a result of a change in the population to which the tests were applied.

A commonly used measure in understanding COVID-19 transmission rate is the moving average. A 7 days moving average for the positive testing rate was also estimated and this gave a similar evolution pattern as the estimated positive testing rate.

From July 19 onward, the change in positive testing rate (the derivative plot presented in upper part of Figure 4) is negative (indicating decline in the positive testing rate) but from August, 11, 2020, the derivative begins to increase. This could suggest a change in the transmission trend and an increasing number of positive cases.

### Discussion

In view of the existing healthcare challenges in South Africa, reliable and accurate knowledge about the positive testing rate of COVID-19 is important to ensure prediction of the disease trajectory, optimal resource allocation and better understanding of the transmission process. In the current study we modelled the COVID 19 cases out of the number of tests as a function of time using semi-parametric approach. This approach allows us to adjust for or take into account the number of tests performed, which when ignored may lead to erroneous conclusions. Also, this method allows us to overcome the problem to modelling the number of cases alone and to take into account the strong relationship between the number of cases and the number of tests which can lead to a misleading result and therefore affect government policy regarding measures and precautions needed.

The positive testing rate decreased from early March when the disease was first observed until early May when it kept on increasing. In July, the infection reached

its peak and then consistently decreased, indicating that the intervention strategy was effective. From mid August, 2020, the rate of change of the positive testing rate indicates the decline in the positive testing rate is slowing down suggesting that a less effective intervention is currently implemented. The moving average is another measure that can be used to understand the rate of infection, but unlike the positive testing rate, the moving average commonly uses partial information since there is always loss of information on both tails. In our case, the same result was obtained using both measures.

The rate of infection can be used as an indicator for the evolution of the outbreak over time and to reveal new trends in the outbreak. One could also extend our approach by modeling jointly the number of tests and number of positive cases. These results need to be interpreted under the background of changing COVID-19 testing strategies in the country. When the positive testing rate is tracked in real time, it can provide useful guidance to policy makers

#### **Acknowledgements**

None.

#### **Funding**

None.

#### **Abbreviations**

WHO : World Health Organisation

SARS: Severe Acute Respiratory Syndrome

MERS: Middle East Respiratory Syndrome

PUI: Persons Under Investigation SARS-COV-2: Severe Acute Respiratory Syndrome Coronavirus 2

SAMRC: South African Medical Research Council

GLMM: Generalized Linear Mixed-Effects Model

LMM: Linear Mixed-Effects Model

#### **Availability of data and materials**

The data analyzed in this study is available on [github](#).

#### **Competing interests**

The authors declare that there are no competing interests.

#### **Consent for publication**

Not applicable.

#### **Ethics approval and consent to participate**

Not applicable. Consent to participate does not apply to our study since it utilized publicly available COVID-19 case data.

#### **Authors' contributions**

SM and OEO are joint first authors of the manuscript and jointly planned the data analysis and wrote the paper. SM and TR were part of the South-Africa team in SAMRC that took part in the study that generated the data. CJ provided materials regarding the methodology. KA, CJ and SZ revised the various drafts for submission. All authors read and approved the final submission of the manuscript.

#### **Author details**

<sup>1</sup>Data Science Institute, Center for Statistics, I-BioStat, Hasselt University, Hasselt, Belgium.. <sup>2</sup>Biostatistics Research Unit, South African Medical Research Council, Capetown, South Africa.. <sup>3</sup>Department of Anthropology, Durham Research Methods Centre, Durham University, Durham, United Kingdom.. <sup>4</sup>School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, KwaZulu-Natal, South Africa.. <sup>5</sup>Department of Statistics, University of Pretoria, Pretoria, South Africa..

#### **References**

1. World Health Organisation: WHO Coronavirus Disease (COVID-19) Dashboard (2020). <https://covid19.who.int/> Accessed [Accessed 12 October 2020]
2. Silal, S., Pulliam, J., Meyer-Rath, G., Nichols, B., Jamieson, L., Kimmie, Z., Moultrie, H.: Estimating cases for covid-19 in south africa on behalf of the south african covid-19 modelling consortium hero 2 (2020)
3. Reddy, T., Shkedy, Z., Rensburg, C., Mwambi, H., Debba, P., Zuma, K., Manda, S.: South africa's trajectory to 100000 cases and what lies ahead: data-driven, real-time prediction of total number of reported cases and deaths. Submitted for publication to BMC Medical research methodology (2020)

4. Chowell, G., Tariq, A., Hyman, J.M.: A novel sub-epidemic modeling framework for short-term forecasting epidemic waves. *BMC Med* **17**, 164 (2019)
5. Hsieh, Y.-H., Chen, C.W.S.: Turning points, reproduction number, and impact of climatological events for multi-wave dengue outbreaks. *Tropical Medicine and International Health* **14**, 628–638 (2009)
6. Sebrango-Rodríguez, C.R., Martínez-Bello, D.A., Sánchez-Valdés, L., Thilakarathne, P.J., Del Fava, E., Van Der Stuyft, P., López-Quílez, A., Shkedy, Z.: Real-time parameter estimation of zika outbreaks using model averaging. *Epidemiology & Infection* **145**, 2313–2323 (2017)
7. Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J.M., Yan, P., Chowell, G.: Short-term forecasts of the covid-19 epidemic in guangdong and zhejiang, china. *Clinical Medicine* **9**(596), 13–23 (2020)
8. Shen, C.Y.: Logistic growth modelling of covid-19 proliferation in china and its international implications. *International Journal of Infectious Diseases* **96**, 582–589 (2020)
9. Tariq, A., Lee, Y., Roosa, K., Blumberg, S., Yan, P., Ma, S., Chowell, G.: Real-time monitoring the transmission potential of covid-19 in singapore. *BMC Medicine* **18**, 166 (2020)
10. Our world in Data: The positive rate: A crucial metric for understanding the pandemic (2020). <https://ourworldindata.org/coronavirus-testing-the-positive-rate-a-crucial-metric-for-understanding-the-pandemic/> Accessed 21 October 2020
11. Liu, R., Han, H., Liu, F., Lv, Z., Wu, K., Liu, Y., Feng, Y., Zhu, C.: Positive rate of rt-pcr detection of sars-cov-2 infection in 4880 cases from one hospital in wuhan, china, from jan to feb 2020. *Clinica chimica acta; international journal of clinical chemistry* **505**, 172–175 (2020)
12. Omori, R., Mizumoto, K., Chowell, G.: Changes in testing rates could mask the novel coronavirus disease (covid-19) growth rate. *international journal of infectious diseases. IJID : official publication of the International Society for Infectious Diseases* **94**, 116–118 (2020)
13. Chiu, W.A., Ndeffo-Mbah, M.L.: Using test positivity and reported case rates to estimate state-level covid-19 prevalence in the united states. <https://www.medrxiv.org/content/10.1101/2020.10.07.20208504v1.article-metrics> (2020)
14. Ruppert, D., Wand, M.P., Carroll, R.J.: *Semi Parametric Regression*. Cambridge University Press, ??? (2003)
15. Claesen1, J., Clement, L., Shkedy, Z., Foulquie-Moreno, M.R., Burzykowski, T.: Simultaneous mapping of multiple gene loci with pooled segregants. *PLoS ONE* **8**, 2 (2013)
16. Wood, S.N.: *Generalized Additive Models: An Introduction with R*, 2nd edn. Chapman and Hall/CRC, ??? (2017)
17. R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2020). R Foundation for Statistical Computing. <https://www.R-project.org/>

#### Figures

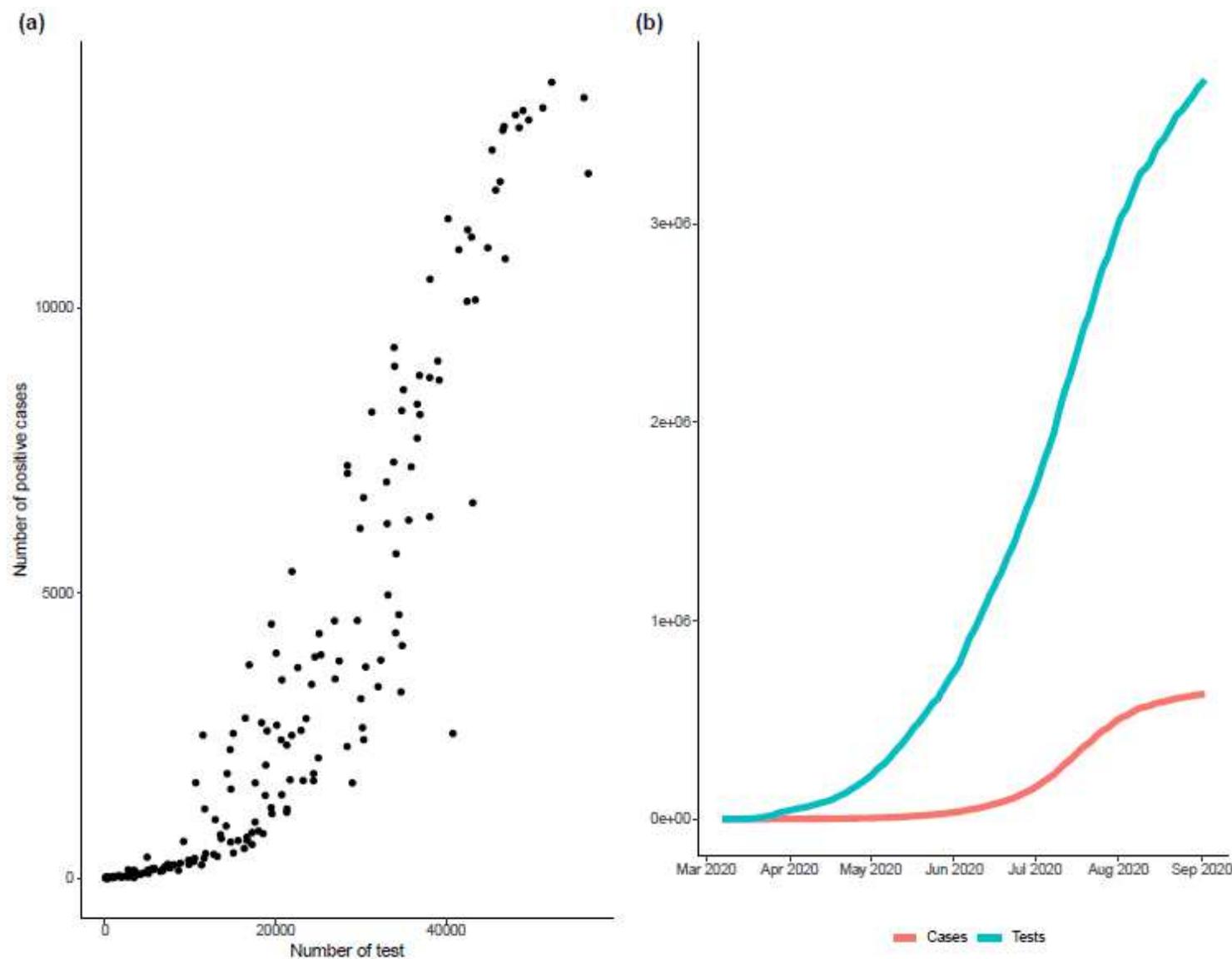
**Figure 1** COVID-19 cases and test over time. Panel a: Relationship between the daily number of COVID-19 tests and the daily number of COVID-19 positive cases. Panel b: The total number of COVID-19 cases and total number of COVID-19 tests carried out between the period March 7th 2020 and September 2nd 2020. Spearman's rank correlation between the number of COVID-19 tests and the number of COVID-19 cases is equal to 0.9632752, ( $p$ -value  $< 2.2e - 16$ ).

**Figure 2** COVID-19 and positive testing rate. Panel a: Daily number of cases and daily number of COVID-19 tests between March 7th, 2020 and September 2nd, 2020. Panel b: Positive testing rate.

**Figure 3** Proportion of infection Observed proportion of infection over time, the estimated probability and the 7 days moving average of positive testing rate.

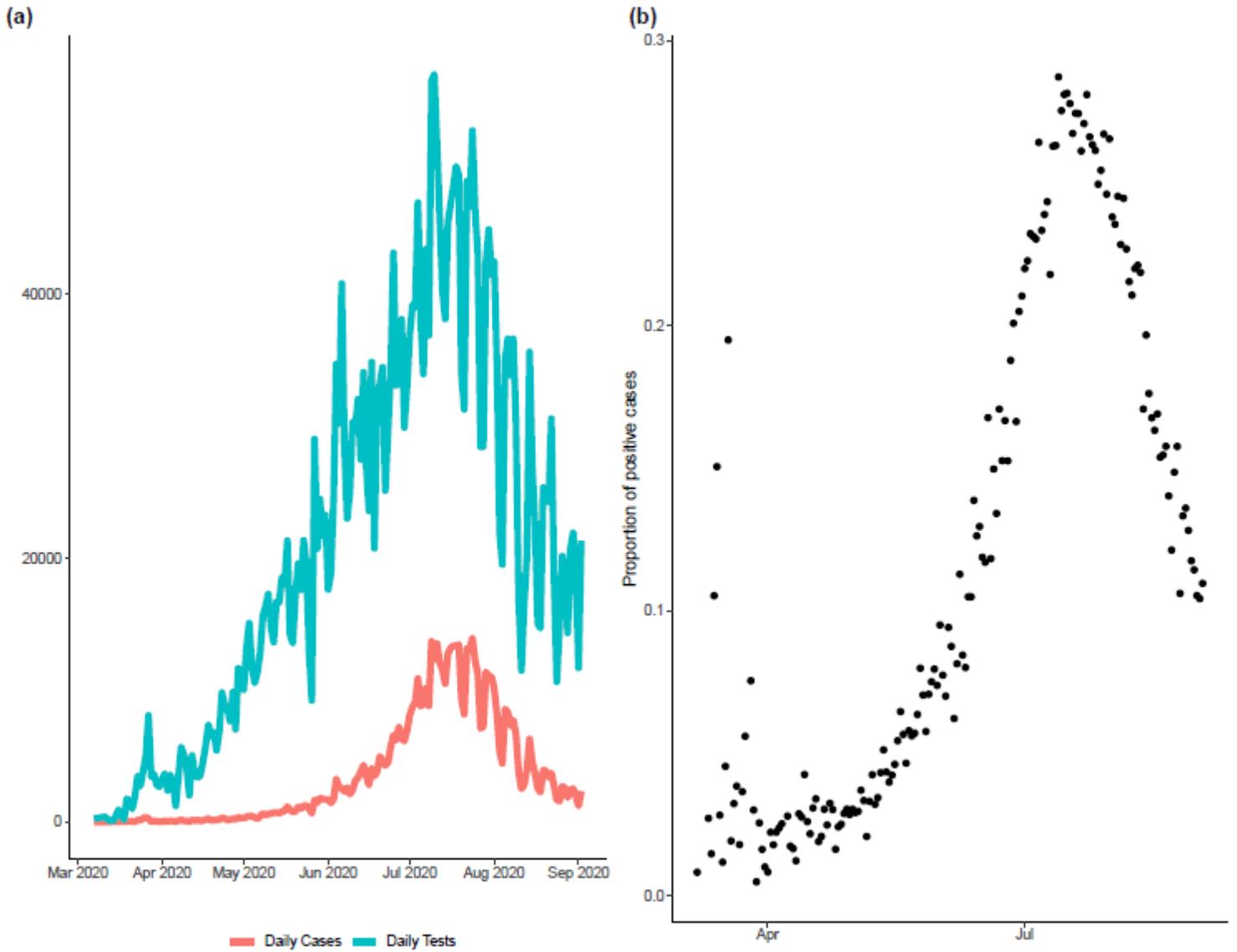
**Figure 4** Graphical Upper panel: Estimated positive testing rate with 95% simultaneous confidence band. Middle panel: The linear predictor of the smoother with 95% simultaneous confidence band. Lower panel: The derivative of the estimated probability with 95% simultaneous confidence band.

# Figures



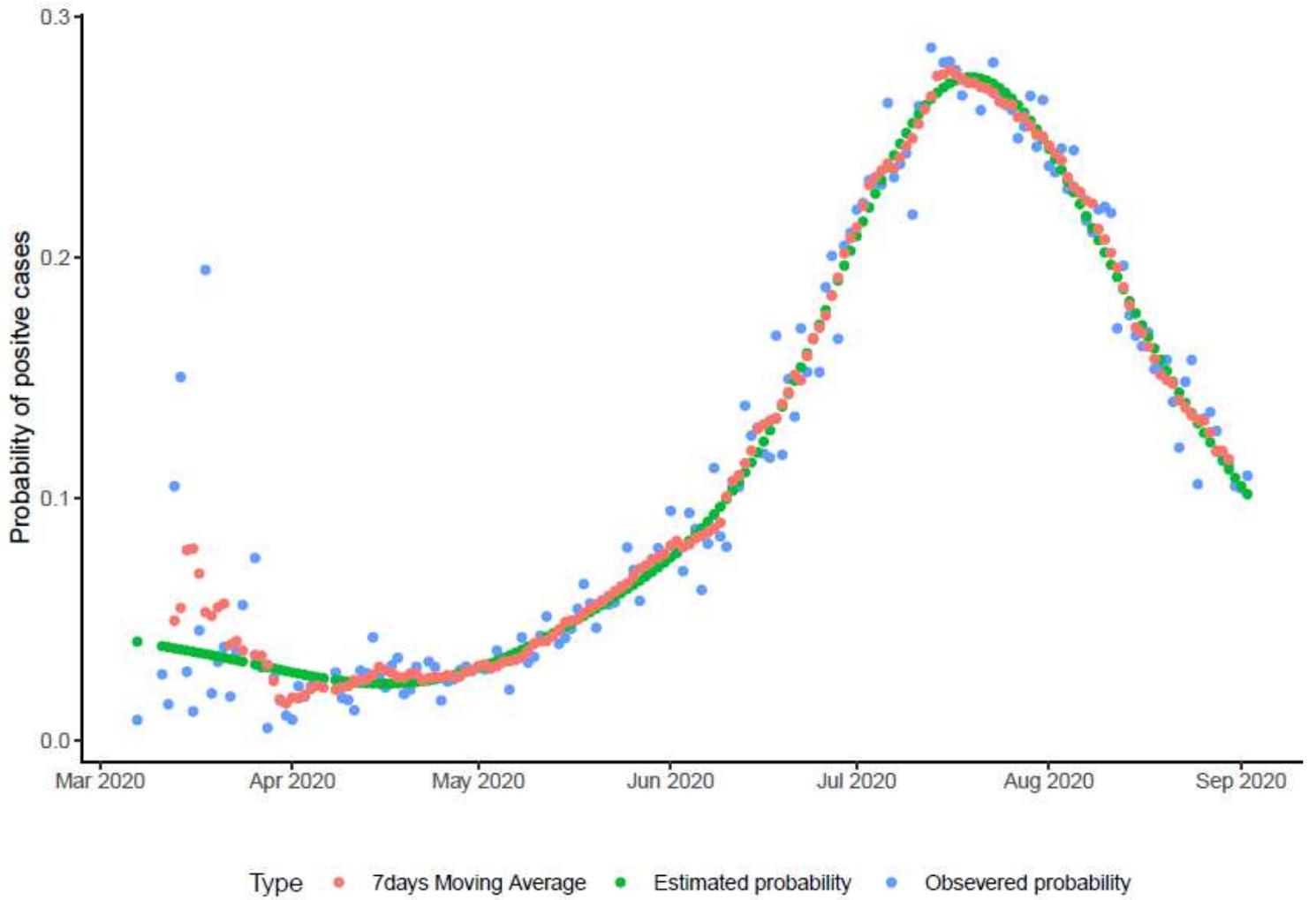
**Figure 1**

COVID-19 cases and test over time. Panel a: Relationship between the daily number of COVID-19 tests and the daily number of COVID-19 positive cases. Panel b: The total number of COVID-19 cases and total number of COVID-19 tests carried out between the period March 7th 2020 and September 2nd 2020. Spearman's rank correlation between the number of COVID-19 tests and the number of COVID-19 cases is equal to 0.9632752, (p-value < 2:2e - 16).



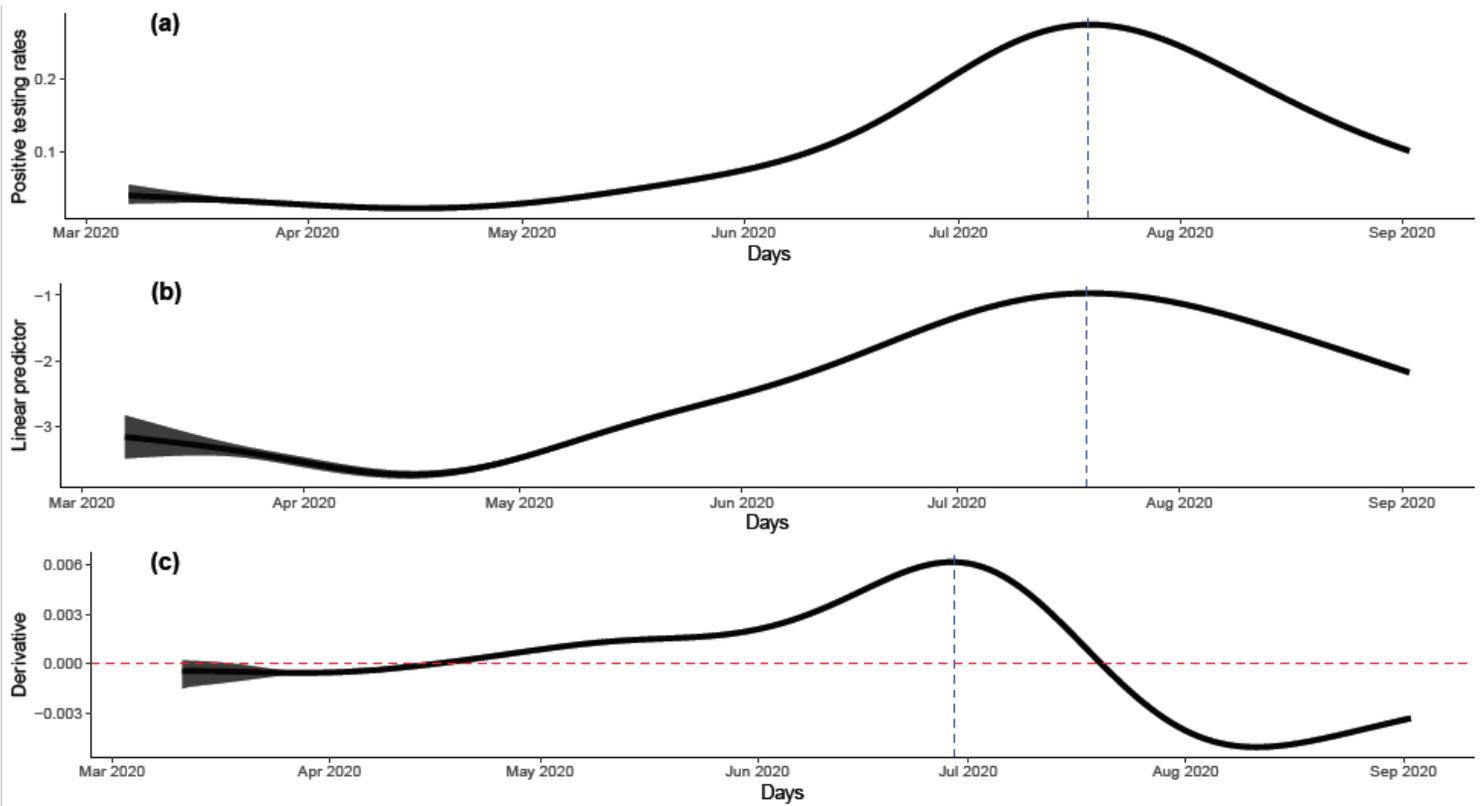
**Figure 2**

COVID-19 and positive testing rate. Panel a: Daily number of cases and daily number of COVID-19 tests between March 7th, 2020 and September 2nd, 2020. Panel b: Positive testing rate.



**Figure 3**

Proportion of infection Observed proportion of infection over time, the estimated probability and the 7 days moving average of positive testing rate.



**Figure 4**

Graphical Upper panel: Estimated positive testing rate with 95% simultaneous condence band. Middle panel: The linear predictor of the smoother with 95% simultaneous condence band. Lower panel: The derivative of the estimated probability with 95% simultaneous condence band.