

## Model-based evaluation of transmissibility and reinfection for the P.1 variant of the SARS-CoV-2 — Source link

Renato Mendes Coutinho, Flavia Maria Darci Marquitti, Leonardo Souto Ferreira, Marcelo Eduardo Borges ...+12 more authors





**Institutions:** Universidade Federal do ABC, State University of Campinas, Spanish National Research Council, Federal University of Maranhão ...+2 more institutions

**Published on:** 05 Mar 2021 - medRxiv (Cold Spring Harbor Laboratory Press)

**Topics:** Transmissibility (vibration)

### Related papers:

- [Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil.](#)
- [Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence.](#)
- [Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 \(SARS-CoV-2\) lineage with multiple spike mutations in South Africa](#)
- [Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England.](#)
- [Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/model-based-evaluation-of-transmissibility-and-reinfection-4bctdilzjr>

# Model-based estimation of transmissibility and reinfection of SARS-CoV-2 P.1 variant

Renato Mendes Coutinho<sup>a,h,1</sup>, Flavia Maria Darcie Marquitti<sup>b,h</sup>, Leonardo Souto Ferreira<sup>c,h</sup>, Marcelo Eduardo Borges<sup>h</sup>, Rafael Lopes Paixão da Silva<sup>c,h</sup>, Otavio Canton<sup>c,h</sup>, Tatiana P. Portella<sup>d,h</sup>, Silas Poloni<sup>c,h</sup>, Caroline Franco<sup>c,h</sup>, Mateusz M. Plucinski<sup>g</sup>, Fernanda C. Lessa<sup>g</sup>, Antônio Augusto Moura da Silva<sup>e,h</sup>, Roberto Andre Kraenkel<sup>c,h</sup>, Maria Amélia de Sousa Mascena Veras<sup>f,h</sup>, and Paulo Inácio Prado<sup>d,h</sup>

<sup>a</sup>Universidade Federal do ABC, Santo André, SP, Brazil; <sup>b</sup>Universidade Estadual de Campinas, Campinas, SP, Brazil; <sup>c</sup>Universidade Estadual Paulista, São Paulo, SP, Brazil; <sup>d</sup>Universidade de São Paulo, São Paulo, SP, Brazil; <sup>e</sup>Universidade Federal do Maranhão, São Luís, MA, Brazil; <sup>f</sup>Faculdade de Ciências Médicas da Santa Casa de São Paulo – São Paulo, SP, Brazil; <sup>g</sup>Centers for Disease Control and Prevention, Atlanta, USA; <sup>h</sup>Observatório COVID-19 BR

This manuscript was compiled on March 17, 2021

1 The variant of concern (VOC) P.1 emerged in the Amazonas state  
2 (Brazil) in November-2020. It contains a constellation of mutations,  
3 ten of them in the spike protein. Consequences of these specific muta-  
4 tions at the population level have been little studied so far, despite  
5 the detection of P.1 variant in 26 countries, with local transmission  
6 in at least four other countries in the Americas and Europe. Here, we  
7 estimate P.1's transmissibility and reinfection using a model-based  
8 approach, by fitting data from the Brazilian national health surveil-  
9 lance of hospitalized individuals and frequency of the P.1 variant in  
10 Manaus from December 2020 to February 2021, when the city was  
11 devastated by four times more cases than in the previous peak (April  
12 2020). The new variant was found to be about 2.6 times more trans-  
13 missible (95% Confidence Interval (CI): 2.4–2.8) than previous circu-  
14 lating variant(s). The city already had a high prevalence of individu-  
15 als previously affected by the SARS-CoV-2 virus (estimated as 78%,  
16 CI:73–83%), and the fitted model attributed 28% of the cases during  
17 the period to reinfections by the variant P.1. Our estimates rank P.1  
18 as the most transmissible among the current identified SARS-CoV-  
19 2 VOCs, posing a serious threat and requiring urgent measures to  
20 control its global spread.

variant of concern | compartmental model | epidemic dynamics

1 **T**he Japanese National Institute of Infectious Diseases iden-  
2 tified the new P.1 SARS-CoV-2 variant from travelers  
3 returning from Amazonas State, Brazil, on 6-January-2021  
4 (1). P.1 was eventually reported in Manaus city (Amazonas  
5 state capital), on 11-January-2021 (2). Later, it was identi-  
6 fied in samples collected since 6-Dec-2020 from Manaus (3).  
7 According to phylogenetic studies, P.1 likely emerged in the  
8 Amazonas state in early (3) or late (4) November 2020. This  
9 variant shares mutations with other variants of concern (VOCs)  
10 previously detected in the United Kingdom and South Africa  
11 (B.1.1.7 and B.1.351, respectively) (2). Mutations of these  
12 two other variants are associated with greater transmissibility  
13 and immune evasion (5, 6), which confer them the status of  
14 variant of concern. However information, data, and analyzes  
15 on the epidemiology of P.1 are still incipient.

16 The Coronavirus disease 2019 (COVID-19) outbreak in  
17 Manaus (April–May 2020) was followed by a period of high  
18 but stable incidence, after which prevalence may have reached  
19 42% (7) to 76% (8) by November 2020. From December 2020  
20 to February 2021 the city was devastated by a new outbreak  
21 that caused a collapse in the already fragile health system  
22 (9), with shortages of oxygen supply (10), while the frequency  
23 of P.1 increased sharply from 0% in November 2020 to 73%

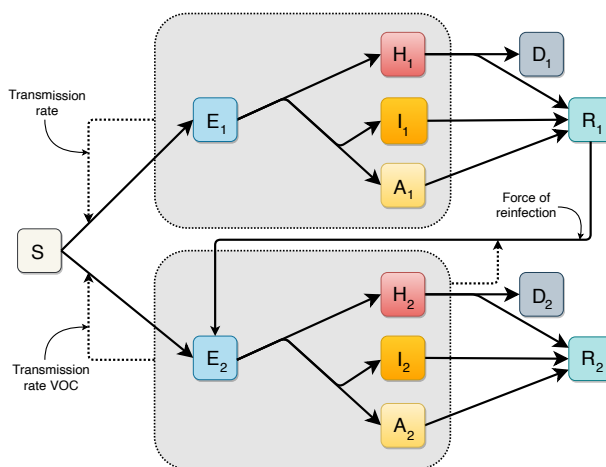


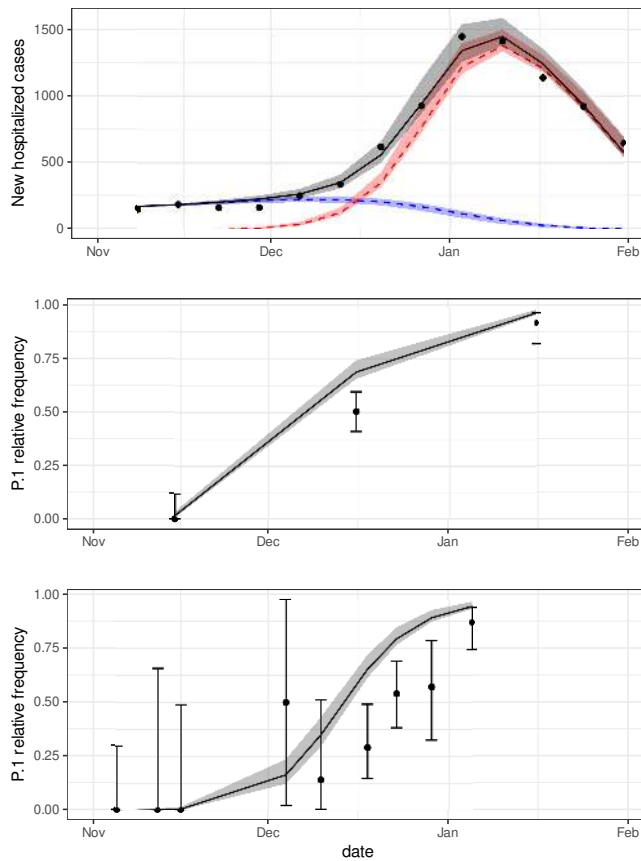
Fig. 1. Diagram of the extended deterministic compartmental model (SEAIHRD). S: Susceptible, E: Exposed (pre-symptomatic), H: Hospitalized (severe infected individuals), I: Infected (symptomatic individuals, not hospitalized), A: Asymptomatic. D: Deceased, R: Recovered. Compartments are subdivided into 3 age categories, not represented here for simplicity. Compartments with subindex 1 represent the wild-type variant, subindex 2 refers to the VOC P.1. Continuous lines represent flux between each compartment; dashed lines, infection pathways. Small arrows indicate force of reinfection and transmissibility.

in January 2021 (4). The pathogenicity of P.1 variant is still  
unknown, although recent studies point to increased viral load  
in individuals infected with the new variant (4), suggesting  
it could be higher than the one from previous circulating  
strain. We analyzed Brazilian national health surveillance  
data on COVID-19 hospitalizations and the frequency of P.1  
among sequences from residents of Manaus city using a model-  
based approach (an extended SEIR compartmental model –  
see Fig. 1) to estimate the transmissibility and relative force  
of reinfection of the P.1 variant.

**Results.** The estimated transmissibility of P.1 was 2.6 (95%  
Confidence Interval (CI): 2.4–2.8) times higher compared to  
the wild-type variant, while the relative force of reinfection  
of the new variant was estimated to be 0.032 (CI: 0.026–  
0.040, Table 1) The fitted model also estimated that, at the

RMC, FMDM, LSF, MEB, RLPS, TPP, SP, CF, RAK, MASMV, and PIP designed research; RMC, FMDM, LSF, MEB, RLPS, TPP, SP, and PIP performed research and analyzed data; and all authors wrote the paper.

<sup>1</sup>To whom correspondence should be addressed. E-mail: renato.coutinhofabc.edu.br



**Fig. 2.** First panel: Weekly new hospitalized COVID-19 cases in Manaus city. Grey line represents the fitted values of total cases (all variants) by maximum likelihood estimation (MLE) of the parameters. Red and blue lines represent the predicted values of cases due to P.1 and wild-type variants, respectively. Black dots are nowcasted observed data of hospitalizations. Second and third panels show the fittings to the time-series frequency of P.1 on datasets provided by (11) and (3) respectively. The area around the lines indicate the 95% Confidence Interval (CI) of the expected values. Dots and lines are the sample proportions of P.1 in sequenced genomes, and their 95% sample CI. The fitted values of the model parameters are presented in Table 1.

time P.1 variant emerged, the prevalence of previous infection by the wild-type variant was 78% (CI: 73–83%), and that the number of cases by the wild-type variant were increasing with an estimated daily intrinsic growth rate of  $0.029 \text{ days}^{-1}$  (CI:  $0.024\text{--}0.035 \text{ days}^{-1}$ ). Given these parameter values, reinfections by P.1 accounted for 28% of the cases in Manaus from November 2020 through January 2021.

We also evaluated the impact of a distinct pathogenicity of the P.1 variant on our estimates by allowing the infection hospitalization rate (IHR) of the new variant to be estimated as a free parameter (see SA1 in Table 1). The relative transmissibility and prevalence did not differ statistically from the the previous estimates. However, the data gives no support for a higher IHR of P.1. Moreover, the model fit to hospitalization data prior to the healthcare system collapse in the city of Manaus (11 January, 2021) estimated an even larger transmissibility (SA2 in Table 1).

**Discussion.** COVID-19 hospitalizations and frequency of the P.1 variant in clinical samples showed a sharp increase in Manaus, Brazil, starting December 2020. The fitted model describes this joint increase as the result of the emergence of

P.1, estimated to be 2.6 times more transmissible than the wild-type variant. The spread of P.1 occurred despite a high estimated prevalence of infection by the wild-type virus. The pathogenicity of P.1 is still unknown, but assuming hospitalization rates as a proxy for pathogenicity, P.1 transmissibility holds for different ranges of pathogenicity. Two recent studies analysed genomic data of SARS-CoV-2 from Manaus evaluating the transmissibility of the new variant (3, 4). Faria and collaborators integrated mortality and genomic data and, using a semi-mechanistic Bayesian model, estimated a transmissibility 1.4–2.2 times higher and 25–61% evasion of protective immunity related to the variant P.1 (3). Naveca and collaborators estimated a 2.2 times higher effective reproduction number for the P.1 variant using phylogenetic methods, and suggested that P.1 is at least two times more transmissible than the parental lineage, assuming reinfections are rare (4). The present work follows a different approach that can be defined as an epidemiological, model-based, and data-fitting approach, suitable for scenarios where only surveillance data are available, and applicable to other emerging variants throughout the world. Notably, all three different approaches estimated very high transmissibility of the P.1 variant.

Many knowledge gaps about the pandemic in the Amazonian region still remain. Population-based serological surveys are not available and thus prevalence was included in the fitted parameters. The analysed data overlapped with the period of the health system collapse. Aware that in-hospital fatality rates can quickly change when the health system is under stress (12), we have chosen hospitalization data instead of mortality data (See **Dataset**). Still, during the health system collapse many severe cases probably were not recorded in the system and remained unaccounted for. Our results were robust to removing this period in the sensitivity analysis (SA2). Even without P.1 emergence, the model estimates an increase in the number of cases (parameter  $r$ , see Table 1), which could be a consequence of loosening non-pharmacological interventions (NPIs) (9), an effect of waning immunity, or both. Our model does not consider these effects explicitly, but by fitting the initial growth rate we indirectly account for their effects on the dynamics and on the estimation of the remaining parameters.

The impacts of a highly transmissible variant have already been highlighted by the spread of VOC B.1.1.7 in the UK, USA and Europe (13). The variant B.1.1.7 has an upper-bound estimate for transmissibility of 2.3 (5), which is smaller than our lower bound estimate for P.1. Higher transmissibility of the P.1 variant raises strong concerns of swift upsurges in the number of cases once P.1 reaches community transmission. Although our estimate for the relative force of reinfection by the variant P.1 seems low, the impact is strong enough to drive, together with a high transmissibility, a large surge even in a population heavily affected by the wild-type variant. For instance, in Manaus, 28% of the new cases in the period considered were due to reinfections by P.1 in our estimations, reaching 40% when assuming a different IHR for P.1 (SA1). However, in a scenario of low prevalence rate of infection by the wild-type variant, the high transmissibility is the most determinant parameter of the rapid increase in the number of cases and can lead to even sharper outbreaks. The P.1 variant has already been detected in at least 26 countries, with local transmission currently confirmed in four of them (13). This points to the urgency of reinforcing measures to avoid a

**Table 1. Summary of the fitted parameters and respective confidence intervals considering the entire period, November-1, 2020–January-31, 2021 maintaining the same pathogenicity of the previous variant. Sensitivity analyses were performed considering different pathogenicity of the P.1 variant (SA1) and data censoring after the collapse of the healthcare system (SA2) in Manaus, Brazil, on 10-Jan-2021.**

\* parameter was fixed, not estimated, in this analysis

Parameter	main fitting			SA1			SA2		
	estimate	2.5 %	97.5%	estimate	2.5 %	97.5 %	estimate	2.5 %	97.5 %
Relative transmission rate for the new variant	2.61	2.45	2.76	2.52	2.28	2.76	2.95	2.70	3.20
Relative force of reinfection of P.1	0.032	0.026	0.040	0.053	0.044	0.065	0.000	0.000	0.000
Prevalence of previous infection (2020-11-01) (%)	78	73	83	73	67	78	71	69	74
Initial fraction of the new variant (2020-11-01) ( $\times 10^{-5}$ )	30.4	8.2	112.9	8.5	1.4	50.8	17.6	5.0	62.4
Intrinsic growth rate ( $\text{days}^{-1}$ )	0.029	0.024	0.035	0.045	0.037	0.052	0.030	0.026	0.034
Relative IHR odds ratio	1*	–	–	0.74	0.63	0.85	1*	–	–

121 global spread of P.1, which include an agile global genomic  
122 surveillance network. Further, to improve our ability to deal  
123 with the threat of P.1, it is urgent to study i) the pathogenicity  
124 of the P.1 variant, since this trait, in association with high  
125 transmissibility, can drive even well-prepared health systems  
126 to collapse; ii) the efficacy of current vaccines for P.1 variant  
127 infections; and iii) the main factors promoting the emergence  
128 of VOCs, specially the roles of previous high prevalence and  
129 of waning immunity.

## 130 Materials and Methods

131  
132 **Dataset.** We used the Brazilian epidemiological syndrome surveil-  
133 lance system for influenza, SIVEP-Gripe (<https://opendatasus.saude.gov.br>),  
134 to track COVID-19 hospitalized cases. All hospitalized  
135 patients with Severe Acute Respiratory Illness (SARI) are reported  
136 to SIVEP-Gripe with symptom onset date and SARS-CoV-2 test  
137 results. SIVEP-Gripe, due to its universal coverage and mandatory  
138 notification of SARI cases, has an homogeneous testing effort to  
139 diagnose SARS-CoV-2 infections, and is currently the best source  
140 for Brazilian data at the national level. Hospitalization data pro-  
141 vides the most accurate basis to infer incidence in Manaus, because  
142 mild cases are vastly under-reported and testing capacity fluctuates,  
143 while mortality data is harder to relate to total number of cases,  
144 since the city's health system endured a prolonged stress even before  
145 the collapse, with large variation in the in-hospital fatality rate over  
146 time (12). Data for hospitalized COVID-19 cases among residents  
147 in Manaus from 01-Nov-2020 to 31-Jan-2021 was obtained from  
148 SIVEP-Gripe database of 15-Feb-2021. The hospitalized cases of the  
149 last 10 weeks in the time series were nowcasted (14) to correct for  
150 notification delay. Time-series of frequency of sequenced genomes  
151 identified as P.1 in Manaus were extracted from published datasets  
152 (3, 11).

153 **Model.** A deterministic compartmental model (Figure 1) was devel-  
154 oped to model the infection of Susceptible individuals moving to the  
155 Exposed (pre-symptomatic) compartment, which can progress to the  
156 three alternative compartments: Hospitalized (severely ill), Infected  
157 (symptomatic but non-hospitalized), and Asymptomatic. Eventu-  
158 ally, individuals move to Recovered or Deceased. Two variants are  
159 considered: 1-non-P.1 (“wild-type”) and 2-new/P.1. The latter is  
160 assumed to infect Recovered individuals previously infected by the  
161 wild-type, and no reinfections of wild-type due to waning immu-  
162 nity occur. Compartments were stratified into three age categories:  
163 young (< 20 years old), adult ( $\geq 20$  and < 60 years old) and elderly  
164 ( $\geq 60$  years old), with different rates for outcomes. The key pa-  
165 rameters of relative transmissibility and relative force of reinfection  
166 – the ratio between the force of infection by P.1 on previously in-  
167 fected individuals (reinfections) and the force of infection by P.1 on  
168 susceptible ones (new infections) – were estimated by a maximum  
169 likelihood fitting to the weekly number of new hospitalizations and  
170 to genomic surveillance data. Three additional model parameters  
171 with unknown values were estimated. The remaining parameters  
172 (24 out of 29) were fixed, using current values from the literature

(see Supplementary Information (SI) for values and references). Sen-  
173 sitivity to different pathogenicity of the variant P.1 was explored  
174 by repeating the fit assuming IHR as a free parameter (SA1). The  
175 sensitivity to the period analysed was also explored by another fit  
176 excluding the health system collapse period (SA2). Further model  
177 details and fitting methodology are available in the SI.  
178

179 **ACKNOWLEDGMENTS.** We are grateful for the collaborative  
180 work of the entire group of the Observatório COVID-19 BR. In par-  
181 ticular, we thank Verônica Coelho for critical inputs. The authors  
182 also thank the research funding agencies: the Coordenação de Aper-  
183 feiçoamento de Pessoal de Nível Superior – Brazil (Finance Code 001  
184 to FMDM, LSF and TPP), Conselho Nacional de Desenvolvimento  
185 Científico e Tecnológico – Brazil (grant number: 315854/2020-0 to  
186 MEB, 141698/2018-7 to RLPs, 313055/2020-3 to PIP, 312559/2020-  
187 8 to MASMV, 311832/2017-2 to RAK, 305703/2019-6 to AAMS)  
188 and Fundação de Amparo à Pesquisa do Estado de São Paulo - Brazil  
189 (grant number: 2019/26310-2 and 2017/26770-8 to CF, 2018/26512-  
190 1 to OC, 2018/24037-4 to SP and contract number: 2016/01343-7  
191 to RAK). **Disclaimer** The findings and conclusions in this article  
192 are those of the authors and do not necessarily represent the official  
193 position of the Centers of Disease Control and Prevention.  
194

- 195 1. NIID Japan, Brief report: New variant strain of sars-cov-2 identified in travel-  
196 ers from brazil. coronavirus disease 4 (2021) <https://www.niid.go.jp/niid/en/2019-ncov-e/10108-covid19-33-en.html>, Accessed on 2021-02-28.
- 197 2. F Naveca, et al., Phylogenetic relationship of sars-cov-2 sequences from amazonas with  
198 emerging brazilian variants harboring mutations e484k and n501y in the spike protein.  
199 [Virological.org](https://virological.org) (2021) Available at: <https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585>.
- 200 3. NR Faria, et al., Genomics and epidemiology of a novel sars-cov-2 lineage in Manaus, Brazil.  
201 [medRxiv](https://medrxiv.org/abs/2021.01.14.21014444), 1–44 (2021).
- 202 4. F Naveca, et al., COVID-19 epidemic in the brazilian state of amazonas was driven by long-  
203 term persistence of endemic SARS-CoV-2 lineages and the recent emergence of the new  
204 variant of concern p.1. *Res. Sq.*, 1–21 (2021).
- 205 5. NG Davies, et al., Estimated transmissibility and impact of sars-cov-2 lineage b.1.1.7 in eng-  
206 land. *Science* (2021).
- 207 6. C Pearson, T Russell, N Davies, et al, Estimates of severity and transmissibility of  
208 novel south africa sars-cov-2 variant 501y.v2 (2021) <https://cmid.github.io/topics/covid19/sa-novel-variant.html>, Accessed on 2021-03-10.
- 209 7. P Lalwani, et al., Sars-cov-2 seroprevalence and associated factors in manaus, brazil:  
210 baseline results from the detectcov-19 cohort study. *SSRN* (2021) Available at SSRN:  
211 <http://dx.doi.org/10.2139/ssrn.3795816>.
- 212 8. LF Buss, et al., Three-quarters attack rate of sars-cov-2 in the brazilian amazon during a  
213 largely unmitigated epidemic. *Science* **371**, 288–292 (2021).
- 214 9. L Ferrante, et al., Brazil's policies condemn amazonia to a second wave of covid-19.  
215 *Nat. Medicine* **26**, 1315–1315 (2020).
- 216 10. L Taylor, Covid-19: Is manaus the final nail in the coffin for natural herd immunity? *bmj* **372**  
217 (2021).
- 218 11. Rede Genômica Fiocruz, Plots of lineages presence by state (2021) <http://www.genomahciv.fiocruz.br/presenca-das-linhagens-por-estado>, Accessed on 2021-02-28.
- 219 12. TP Portella, et al., Temporal and geographical variation of covid-19 in-hospital fatality rate in  
220 brazil. [medRxiv](https://medrxiv.org/abs/2021.03.04.21014444) (2021).
- 221 13. PANGO lineages, Global report investigating novel coronavirus haplotypes – grinch (2021)  
222 [https://cov-lineages.org/global\\_report\\_P.1.html](https://cov-lineages.org/global_report_P.1.html), Accessed on 2021-03-04.
- 223 14. SF McGough, MA Johansson, M Lipsitch, NA Menzies, Nowcasting by bayesian smooth-  
224 ing: A flexible, generalizable model for real-time epidemic tracking. *PLoS Comput. Biol.* **16**,  
225 e1007735 (2020).

# 1 **Supplementary Information for**

## 2 **Model-based estimation of transmissibility and reinfection of SARS-CoV-2 P.1 variant**

3 **Renato Mendes Coutinho, Flavia Maria Darcie Marquitti, Leonardo Souto Ferreira, Marcelo Eduardo Borges, Rafael Lopes**  
4 **Paixão da Silva, Otavio Canton, Tatiana P. Portella, Silas Poloni, Caroline Franco, Mateusz M. Plucinski, Fernanda C. Lessa,**  
5 **Antônio Augusto Moura da Silva, Roberto Andre Kraenkel, Maria Amélia de Sousa Mascena Veras and Paulo Inácio Prado**

6 **RMC, E-mail: [renato.coutinho@ufabc.edu.br](mailto:renato.coutinho@ufabc.edu.br)**

### 7 **This PDF file includes:**

- 8     Supplementary text
- 9     SI References



## 10 Supporting Information Text

11 In order to estimate key parameters of the variant of concern (VOC) P.1, we developed a model and fitted it to time-series  
 12 data of number of hospitalized cases and frequency of the P.1 variation. The fitting approach used here can be applied to other  
 13 regions where data is relatively scarce. It primarily requires weekly incidence data to determine proper model initial conditions.  
 14 In Brazil, these are the hospitalized cases data. Stratification by age allows the model to also consider the different death rates,  
 15 asymptomatic and hospitalized proportions of each age class, important features for SARS-CoV-2. Contact levels between  
 16 different age classes, which may vary from one place to another, can also be considered. For special cases in which information  
 17 such as contact between ages classes and age distribution are not available (or even unnecessary for some other disease), the  
 18 model can be easily simplified. In this sense, the method proposed here demands low-detailed data and relies on the structure  
 19 of a simple compartmental model to measure quantities of interest, such as transmissibility and relative force of reinfection.

20 Section 1 describes the model, section 2 relates the values of the parameters taken from the current literature, section 3  
 21 describes the contact matrix used, and finally section 4 describes the treatment of case data (subsec. A), the choice of initial  
 22 conditions (subsec. B), the fitting procedure (subsec. C), and the sensitivity analysis evaluated regarding pathogenicity and  
 23 data period analysed (subsec. D)

### 24 1. Model equations

25 The model is an extended Susceptible, Exposed, Infected, and Recovered (SEIR) model that comprises susceptible (S),  
 26 pre-symptomatic (E), asymptomatic (A), mild symptomatic (I), severe/hospitalized (H), recovered (R) and deceased (D)  
 27 compartments. These compartments are duplicated to account for a second variant of SARS-CoV-2, and each of them is  
 28 stratified into three age classes: young (<20 years old), adults ([20 – 59] years old), and the elderly ( $\geq 60$  years old). The  
 29 “wild-type” classes represent all non-P.1 variants present, which do not seem to be variants of concern.

30 We assume that the second variant is capable of reinfecting individuals who have recovered from infection by the wild-type  
 31 variant while the inverse is not possible; in the absence of data indicating this possibility, allowing reinfection by the wild-type  
 32 variant on recovered of infection by P.1 would have negligible effect due to the small time window (3 months) considered in the  
 33 present work. We also consider that a variant is not capable of reinfecting individuals recovered from the same lineage. Our  
 34 model does not include vaccination due to low rates of vaccination in Brazil during the study time period.

To model the virus spread in the population, we assume that asymptomatic individuals have equal infectiousness compared  
 to symptomatic ones, while pre-symptomatic individuals have reduced infectiousness represented by  $\omega$ . To model behaviour,  
 we assume that symptomatic individuals self-isolate themselves to some degree, reducing their contacts by  $\xi$ . Individuals with  
 severe disease have greater isolation  $\xi_{sev}$  due to hospitalization. The daily contacts between each age class is represented by  
 the matrix  $\hat{C}$ . The force of infection  $\lambda_k$  for each variant  $k$  is defined below:

$$\lambda_k = \beta_k \hat{C} [A_k + \omega E_k + (1 - \xi)I_k + (1 - \xi_{sev})H_k]$$

The complete system of equations is given by:

#### Completely Susceptible

$$\frac{dS}{dt} = -\lambda_1 \frac{S}{N} - \lambda_2 \frac{S}{N} \quad [1a]$$

#### Wild variant

$$\frac{dE_1}{dt} = \lambda_1 \frac{S}{N} - \frac{E_1}{\gamma_1} \quad [1b]$$

$$\frac{dA_1}{dt} = \frac{(1 - \sigma_1)\alpha_1 E_1}{\gamma_1} - \frac{A_1}{\nu_{i,1}} \quad [1c]$$

$$\frac{dI_1}{dt} = \frac{(1 - \alpha_1)(1 - \sigma_1)E_1}{\gamma_1} - \frac{I_1}{\nu_{i,1}} \quad [1d]$$

$$\frac{dH_1}{dt} = \frac{\sigma_1 E_1}{\gamma_1} - \frac{H_1}{\nu_{s,1}} \quad [1e]$$

$$\frac{dR_1}{dt} = \frac{A_1}{\nu_{i,1}} + \frac{I_1}{\nu_{i,1}} + \frac{(1 - \mu_1)H_1}{\nu_{s,1}} - p_r \lambda_2 \frac{R_1}{N} \quad [1f]$$

$$\frac{dD_1}{dt} = \frac{\mu_1 H_1}{\nu_{s,1}} \quad [1g]$$

## P.1 variant

$$\frac{dE_2}{dt} = \lambda_2 \frac{S}{N} - \frac{E_2}{\gamma_2} + p_r \lambda_2 \frac{R_1}{N} \quad [1h]$$

$$\frac{dA_2}{dt} = \frac{(1 - \sigma_2)\alpha_2 E_2}{\gamma_2} - \frac{A_2}{\nu_{i,2}} \quad [1i]$$

$$\frac{dI_2}{dt} = \frac{(1 - \alpha_2)(1 - \sigma_2)E_2}{\gamma_2} - \frac{I_2}{\nu_{i,2}} \quad [1j]$$

$$\frac{dH_2}{dt} = \frac{\sigma_2 E_2}{\gamma_2} - \frac{H_2}{\nu_{s,2}} \quad [1k]$$

$$\frac{dR_2}{dt} = \frac{A_2}{\nu_{i,2}} + \frac{I_2}{\nu_{i,2}} + \frac{(1 - \mu_2)H_2}{\nu_{s,2}} \quad [1l]$$

$$\frac{dD_2}{dt} = \frac{\mu_2 H_2}{\nu_{s,2}} \quad [1m]$$

## Supplementary Equations

$$C_1(t) = \int_0^t \chi \sigma_1 \frac{E_1(t')}{\gamma_1} dt' \quad [1n]$$

$$C_2(t) = \int_0^t \chi \sigma_2 \frac{E_2(t')}{\gamma_2} dt', \quad [1o]$$

where  $C_1$  and  $C_2$  are the cumulative hospitalization cases reported, and each variable of the system ( $S, E_k, \dots, C_k$ ) is a vector containing each age class, e.g.,  $E_1 = (E_{1,y}, E_{1,a}, E_{1,e})^T$ . The equations were numerically solved by the R package developed by (1).

## 2. Parameterization of the model

The parameters considered for the wild-variant are described below. The parameters for the P.1 variant are the same except for those considered in the model fitting.

- $\gamma$ , Average time in days between being infected and developing symptoms: 5.8 (2)
- $\nu_i$ , Average time in days between being infectious and recovering for asymptomatic and mild cases: 9.0 (3)
- $\nu_s$ , Average time between being infectious and recovering/dying for severe cases: 8.4 SIVEP-Gripe for São Paulo State
- $\xi$ , reduction on the exposure of symptomatic cases (due to symptoms/quarantining): 0.1 [Assumed]
- $\xi_{sev}$ , Reduction on the exposure of severe cases (due to hospitalization): 0.9 [Assumed]
- $\omega$ , Relative infectiousness of pre-symptomatic individuals: 1.0 [Assumed]
- $\alpha$ , Proportion of asymptomatic cases [0.67,0.44,0.31] for Juvenile (4), Adult and Elderly (5)
- $\sigma$ , Proportion of infections that require hospitalization: [0.001,0.012,0.089]\* (6)
- $\mu$ , In-hospital mortality ratio: [0.417,0.188,0.754] (7)
- $\chi$ , Case report probability: 1.0 [Assumed]

\*The proportion is weighted by the age distribution of the population with each age category.

## 3. Contact Matrices

Our model includes three age group categories: young ([0 – 19] y.o.), adults ([20 – 59] y.o.), and elderly (greater than 60y.o.). To model contacts between these groups we use estimated contact matrices computed by (8), but since the original matrices use five-year age bins going up to 95+ years, we aggregate classes leading to a  $3 \times 3$  matrix in the following way:

Let  $A, B$  be sets of indexes forming age groups (not necessarily of equal sizes),  $x_{i,j}$  denoting contact between age groups  $i$  and  $j$  in the original matrix,  $d_i$  denoting population size of the age group  $i$ . The new contact matrix  $\hat{C}$  is given by:

$$\hat{C}_{A^*, B^*} = \frac{\sum_{i \in A} \sum_{j \in B} d_i x_{i,j}}{\sum_{i \in A} d_i} \quad [2]$$

where  $A^*, B^*$  denotes a new indexation rule. Note that the contact matrices depend on local demographics and therefore must be computed for each place of study.

## 61 4. Data Analysis Procedures

62 **A. Nowcasting.** Data used in parameters estimation were collected from the national public health system of severe acute  
63 respiratory illness (SARI) surveillance database, named *Sistema de Vigilância da Gripe - SIVEP-Gripe*. In this system,  
64 reporting of cases can be delayed for several reasons, including the notification system itself and confirmation of RT-PCR test  
65 results. The nowcasting procedure estimates, based on the past delay distribution, the number of cases that already occurred  
66 but were not yet reported. A window of 10 weeks is the acting window on the series, since delays greater than this are rare.

67 Nowcasting requires a pair of dates: (i) onset date of the event and (ii) report date of the event. The delay distribution is  
68 modeled as being best described as a Poisson distribution for days since the onset date to the report date. We considered *the*  
69 *first symptoms date* as the onset date. For the report date, we used the latest between *the test result date* and *the clinical*  
70 *classification date*. The nowcasting algorithm were developed by (9), and implemented in the **NobBS** (Nowcasting by Bayesian  
71 Smoothing) package in R (10).

72 **B. Initial Condition Estimation.** The model requires appropriate mid-epidemic initial conditions in order to give relevant results.  
73 In the model, the number of new hospitalizations at a given time –  $h_{new}$ , is directly proportional to the number of exposed  
74 individuals at that time, therefore data was used to get an approximation of the number of exposed people. Also, to quantify  
75 the number of people belonging to the recovered class, prevalence was used.

76 We can estimate the appropriate initial conditions by finding an approximation for our model that relates more directly to  
77 the available data in each class. In the absence of the variant P.1, the model has four classes of infected compartments, namely  
78  $\mathbf{y} = (E_1, A_1, I_1, H_1)^T$ , and another three classes, represented by  $\mathbf{z}$ , i.e.,  $\mathbf{z} = (S, R_1, D_1)^T$ . To that effect, we can write the  
79 system as

$$80 \quad \dot{\mathbf{y}} = F(\mathbf{y}, \mathbf{z}) - G(\mathbf{y}, \mathbf{z}), \quad [3]$$

$$81 \quad \dot{\mathbf{z}} = J(\mathbf{y}, \mathbf{z}), \quad [4]$$

82 where  $F$  comprises all entries of new Infected, coming from classes  $\mathbf{z}$ , whilst  $G$  accounts for the transitions within infected  
83 classes and also recovery and death from the disease. Then, to find a good approximation for a small time window, we perform  
84 a linearization of our model around a point  $(\mathbf{y}, \mathbf{z})$ . Keeping  $\mathbf{z}$  fixed, we get

$$85 \quad \dot{\mathbf{y}} = (\hat{F} - \hat{G})\mathbf{y}, \quad [5]$$

86 where  $\hat{F}$  and  $\hat{G}$  are the linearized matrices arising from the functions  $F$  and  $G$ , respectively. The only entrance of new infected  
87 comes from the  $\beta S \lambda / N$  terms in the  $\hat{E}_1 = (\hat{E}_{1,y}, \hat{E}_{1,a}, \hat{E}_{1,e})^T$  equations (sub-indexes are  $y$  young,  $a$  adults and  $e$  elderly), then,  
88 the only non-zero elements of  $\hat{F}$  are in its first 3 lines. Before proceeding, it is useful to define

$$89 \quad \hat{b} = \text{diag}(S)\hat{C} \quad [6]$$

90 which allow us to write

$$91 \quad \hat{F} = \frac{\beta}{N} \begin{bmatrix} \omega \hat{b} & \hat{b} & (1 - \xi) \hat{b} & (1 - \xi_{sev}) \hat{b} \\ & & & \\ & & & \\ & & & \end{bmatrix} \quad [7]$$

92  $\hat{G}$  contains the terms of Exposed,  $E_1$ , the 3 possible forms of the disease considered in the model ( $A_1, I_1$  and  $H_1$ ), as the  
93 terms in its first 3 rows, , whilst the remainder of its main diagonal contains terms of recovery and death. For simplicity, every  
94 constant (or vector for the terms with  $\sigma$ ) in  $\hat{G}$  expression Eq. (8) should be thought as diagonal matrices with its elements  
95 given by the constants (or vectors) and every  $\mathbb{0}$  is a 3-dimensional square matrix where all entries are null.

$$96 \quad \hat{G} = \begin{bmatrix} \gamma^{-1} & 0 & 0 & 0 \\ -\alpha(1 - \sigma)\gamma^{-1} & \nu_i^{-1} & 0 & 0 \\ -(1 - \alpha)(1 - \sigma)\gamma^{-1} & 0 & \nu_i^{-1} & 0 \\ -\sigma\gamma^{-1} & 0 & 0 & \nu_s^{-1} \end{bmatrix} \quad [8]$$

97 The linearization above implies that, for a small time interval,  $\mathbf{y}$  has an exponential behavior and that the eigenvalues  
98 of  $\hat{L} = \hat{F} - \hat{G}$  are related to the exponential growth rates. Therefore, a short time after the beginning of the epidemic, the  
99 largest eigenvalue should be the one to dominate. So the exponential growth rate of the wild-type variant –  $r$ , can be matched  
100 to the largest eigenvalue of  $\hat{L}$  to obtain an estimate for  $\beta$ . The eigenvector associated with the largest eigenvalue gives the  
101 proportions of infected classes, which, together with the estimated number of exposed individuals –  $E_1 = \gamma_1 h_{new} / \sigma_1$ , results in  
102 an approximation for the number of people in the other infected classes.

103 Given a  $\beta$ , the largest eigenvalue of the linearization matrix is computed using the `eigs` function of the R package `rARPACK`  
104 (11) and we find the  $\beta$  that gives  $r$  as the largest eigenvalue through bisection root finding. Finally, subtracting the number of  
105 recovered and infected from the total population gives the number of susceptible individuals.



106 **C. Maximum Likelihood Estimation.** Given the cumulative daily curves of hospitalization for wild-type variant,  $C_1$ , and P.1  
 107 variant,  $C_2$ , we can obtain the daily variation of each curve, namely  $\Delta C_1^t$  and  $\Delta C_2^t$ . Those curves are summed up to give the  
 108 total number of weekly new cases:

$$109 \quad \Delta C^\tau = \sum_{i=1}^7 (\Delta C_1^{\tau-1+i} + \Delta C_2^{\tau-1+i}) \quad [9]$$

110 where  $\tau$  is a discrete index given in weeks.

111 To calculate the frequency of P.1 in a given time period  $T$ , we use the proportion of new cases in this period from the  
 112 wild-type and P.1 variant as follows:

$$113 \quad P^{t'} = \frac{\sum_{i=1}^T \Delta C_2^{T-1+i}}{\sum_{i=1}^T \Delta C_1^{T-1+i} + \sum_{i=1}^T \Delta C_2^{T-1+i}} \quad [10]$$

114 where  $t'$  is a discrete index given in  $T$  periods.

115 The time period  $T$  depends on the dataset of genome sequences: it is daily in (12) and monthly in (13).

116 Using maximum likelihood, we fitted the model by estimating five parameters, namely, the relative transmissibility ( $\Delta\beta = \frac{\beta_2}{\beta_1}$ ),  
 117 the relative force of reinfection of P.1 ( $p_r$ ), initial total prevalence ( $\rho^0 = [R/N]_{t=0}$ ), initial fraction of cases that were caused by  
 118 the new variant ( $P^0$ ), and intrinsic growth rate of the wild-type variant ( $r$ ). The initial fraction of P.1 cases ( $P^0$ ) accounts for  
 119 the uncertainty in the time of emergence of the new variant: the simulation starts at beginning of November, so this initial value  
 120 is below 1 individual, and only reaches this threshold by mid- to late November, depending on the value of  $P^0$ . The parameter  
 121  $r$  incorporates effects related to contact rates for the wild-type variant, such as non-pharmacological interventions relaxation,  
 122 elections, and others; it affects the transmissibilities of both variants ( $\beta_1$  and  $\beta_2$ ) in the same way, and so is independent of  $\Delta\beta$ .

123 Number of hospitalization cases were assumed to follow a Poisson distribution, with expected value given by equation  
 124 Eq. (9). The recorded number of P.1 in genome samples was assumed to follow a binomial distribution with an expected value  
 125 equal to the product of the total number of genome sequences sampled in each date and the proportion of P.1 cases (equation  
 126 Eq. (10)). The log-likelihood function for the model fitting was then:

$$127 \quad \mathcal{L} = \sum_i \log \text{Pois}(x^i | \lambda = C^i) + \sum_j \log \text{Bin}(y^j | N = n^j, \theta(\pi^j) = P^j), \quad [11]$$

128 where  $\text{Pois}$  is a Poisson distribution with parameter  $\lambda$ ,  $x^i$  is the number of recorded hospitalizations in week  $i$ ,  $\text{Bin}$  is a Binomial  
 129 distribution with parameters  $N$  (total number of trials) and  $\pi^j$  (probability of success at each trial),  $n^j$  is total number of  
 130 sequences in clinical samples in week or day  $j$ ,  $y^j$  is the number of P.1 sequences in each of these samples, and  $\theta(\cdot)$  is the logit  
 131 function.

132 The model was then fitted by finding the values of the five above mentioned parameters that minimize the negative of the  
 133 log-likelihood function (equation 11), using the function `mle2`, from the R package `bbmle` (14).

134 To find starting values for the optimization performed by `mle2` we calculated the log-likelihood function for one million  
 135 combinations of parameters values in a regular reticulate within reasonable ranges. The 100 sets of parameters that were local  
 136 minima and with highest log-likelihood were used as starting values for the computational minimization.

137 The confidence intervals for the expected number of cases and frequency were estimated from 20000 parametric bootstrap  
 138 samples assuming that the estimated parameters follow a multivariate normal distribution. The parameters of these multivariate  
 139 distributions were the estimated values and estimated variance-covariance matrix of the parameters. For each sampled  
 140 combination of para 2.5% and 95% quantiles of the distribution of bootstrapped expected values.

141 **D. Sensitivity analysis.** The model fitting assumed a constant infection hospitalization rate (IHR, parameter  $\sigma$ ) for each age  
 142 group over time for both variants. Recent evidence suggests that prior SARS-CoV-2 infection protects most individuals against  
 143 reinfection (15), so reinfections might have lower IHR. Because the pathogenicity of the variant P.1 is unknown, the model  
 144 fitting was repeated assuming that the odds ratio of the IHR in each age class for P.1 infections compared to wild-type variant  
 145 infections (SA1) is a free parameter. Moreover, as the collapse of Manaus health system hindered hospitalizations of new severe  
 146 cases and may have affected case recording in surveillance databases, the model fitting was repeated considering only the  
 147 period prior to the collapse (10-January-2021) (SA2).

## 148 References

- 149 1. K Soetaert, T Petzoldt, RW Setzer, Solving differential equations in R: Package `deSolve`. *J. Stat. Softw.* **33**, 1–25 (2010).
- 150 2. Y Wei, et al., A systematic review and meta-analysis reveals long and dispersive incubation period of covid-19. *medRxiv*  
 151 (2020).
- 152 3. M Cevik, et al., SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and  
 153 infectiousness: a systematic review and meta-analysis. *The Lancet Microbe* **2**, e13–e22 (2021).

- 154 4. SM de Saúde Município de São Paulo, Inquérito sorológico para Sars-Cov-2: Prevalência da infecção em escolares das  
155 redes públicas e privada da cidade de São Paulo ([http://www.capital.sp.gov.br/arquivos/pdf/2021/coletiva\\_saude\\_14-01.pdf](http://www.capital.sp.gov.br/arquivos/pdf/2021/coletiva_saude_14-01.pdf))  
156 (2021) [Online; accessed 31-January-2021].
- 157 5. WW Sun, et al., Epidemiological characteristics of COVID-19 family clustering in Zhejiang Province.  
158 *Chin. journal preventive medicine* **54**, 625–629 (2020).
- 159 6. H Salje, et al., Estimating the burden of SARS-CoV-2 in france. *Science* **369**, 208–211 (2020).
- 160 7. TP Portella, et al., Temporal and geographical variation of COVID-19 in-hospital fatality rate in brazil. *medRxiv* (2021).
- 161 8. K Prem, et al., Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data  
162 for the covid-19 era. *medRxiv* (2020).
- 163 9. SF McGough, MA Johansson, M Lipsitch, NA Menzies, Nowcasting by bayesian smoothing: A flexible, generalizable  
164 model for real-time epidemic tracking. *PLoS Comput. Biol.* **16**, e1007735 (2020).
- 165 10. S McGough, ML Nicolas Menzies, M Johansson, *NobBS: Nowcasting by Bayesian Smoothing*, (2020) R package version  
166 0.1.0.
- 167 11. Y Qiu, J Mei, MY Qiu, *r‘ARPACK’: Solvers for Large Scale Eigenvalue and SVD Problems*, (2016) R package version  
168 0.11-0.
- 169 12. NR Faria, et al., Genomics and epidemiology of a novel sars-cov-2 lineage in Manaus , Brazil. *medRxiv* , 1–44 (2021).
- 170 13. Rede Genômica Fiocruz, Plots of lineages presence by state (2021) <http://www.genomahcov.fiocruz.br/presenca-das-linhagens-por-estado>, Accessed on 2021-02-28.
- 171 14. B Bolker, R Development Core Team, *bbmle: Tools for General Maximum Likelihood Estimation*, (2020) R package  
172 version 1.0.23.1.
- 173 15. VJ Hall, et al., Do antibody positive healthcare workers have lower sars-cov-2 infection rates than antibody negative  
174 healthcare workers? large multi-centre prospective cohort study (the siren study), england: June to november 2020.  
175 *medRxiv* , 1–35 (2020).  
176