

## **Estimating the ascertainment rate of SARS-CoV-2 infection in Wuhan, China: implications for management of the global outbreak**

Theodore Lytras<sup>1</sup>, Georgios Panagiotakopoulos<sup>1</sup>, Sotirios Tsiodras<sup>1,2</sup>

1. National Public Health Organization, Athens, Greece
2. 4th Department of Internal Medicine, Attikon University Hospital, Medical School, National and Kapodistrian University of Athens, Greece

### **Corresponding author:**

Theodore Lytras, MD, PhD

National Public Health Organization, Athens, Greece.

Agrafon 3-5, 15123 Marousi, Athens, Greece

Email: [thlytras@gmail.com](mailto:thlytras@gmail.com)

Word count: 1566 (full text), 72 (abstract)

## **Abstract**

We sought to estimate the ascertainment ratio of SARS-CoV-2 infection in Wuhan, China, using a modified Bayesian SEIR model with publicly reported case data. We estimated it at 0.465% (95%CI: 0.464–0.466%), implying that the outbreak in Wuhan was abated by depletion of susceptibles, rather than public health action alone. This suggests a high-transmissibility/low-severity profile for the current pandemic and raises doubt about whether suppression, rather than mitigation, is a feasible goal.

## **Keywords**

coronavirus; SARS-CoV-2; COVID-19; pandemic; epidemic modelling; Bayesian; Wuhan; depletion of susceptibles

The coronavirus disease 2019 (COVID-19) outbreak originated at the end of 2019 in the city of Wuhan, China, and is currently sweeping the globe, with more than 370,000 confirmed cases and 16,000 deaths. With the outbreak having already turned into the latest global pandemic, one of the key unknowns that will determine its public health impact is the Infection Fatality Rate (IFR), i.e. the number of deaths as a proportion of all persons infected with the SARS-CoV-2 novel coronavirus, the causative pathogen of COVID-19. The IFR is expected to be much lower than the Case Fatality Rate (CFR), the proportion among confirmed cases, if many SARS-CoV-2 infections with mild or no symptoms are not being ascertained [1]. A low ascertainment rate has been previously suggested, based on epidemic modelling of exported cases: between 1.8% and 14% in Wuhan [2–5], 28.4% in Italy [6] and just 0.23% in Iran [7].

After the lockdown that came into force on 23 January 2020, Wuhan approximates a closed population in which the outbreak appears to have run its course, with cases on a steady decline since mid-February. We sought to estimate the ascertainment rate by applying epidemic modelling to publicly reported confirmed COVID-19 cases in Wuhan; a secondary objective was to estimate other epidemic parameters, including the basic reproduction number  $R_0$ .

## **Estimating the ascertainment rate in Wuhan**

We downloaded the most recent confirmed case data from Wuhan using the R package ‘nCov2019’ [8], for the period between 3 January and 9 March 2020. On 12 February there was a spike of 13,436 retroactively reported confirmed cases due to a change in the case definition; we addressed this by linearly interpolating the new confirmed cases for that date and distributing the remainder to all previous dates in proportion to the number of cases already reported.

The median incubation period of COVID-19 has been estimated at approximately 5-6 days, with a range between 1-14 days [9,10]. According to the WHO-China joint mission report, the median time from symptom onset to laboratory confirmation in Wuhan decreased from 15 days (range 10-21) in “early January” to 5 days (3-9) in “early February” [11]. To reflect the above, we modelled the time between case reporting and symptom onset with a Weibull distribution with a shape/scale parameter of 8/15.7 on January 10, decreasing linearly to 4/5.5 on February 10 and after. We similarly modelled the time between case reporting and case exposure (infection with SARS-CoV-2) with a Weibull distribution with shape/scale

4.4/22.3 decreasing linearly to 2.6/12.1 (Figure 1). These distributions were used to reconstruct estimated epidemic curves by infection date and by symptom onset, up until February 23, discarding the last two weeks available.

To these data we subsequently fitted a discrete-time SEIR (Susceptible-Exposed-Infected-Removed) model, modified for underascertainment and fitted in a Bayesian framework. The model assumes that cases are observed from only a fraction of the population (i.e. the ascertainment rate) that is stable over time. We further assumed that cases are isolated upon confirmation and are thus removed from the “Infected” pool at that time. Transition probabilities at time  $t$  between S-E, E-I and I-R are  $1-\exp(-\beta I_{t-1}/N)$ ,  $1-\exp(-\kappa)$  and  $1-\exp(-\gamma)$  respectively, where  $I_{t-1}$  the total number of all infected at time  $t-1$ ,  $N$  the population of Wuhan (11 million people),  $\beta$  the contact rate,  $\gamma$  the removal rate and  $\kappa$  the incubation rate (the inverse of the incubation period). In this model,  $\beta/\gamma$  equals the basic reproduction number  $R_0$  and  $(1/\kappa + 1/2\gamma)$  equals the serial interval [12]. We used non-informative Uniform(0,10) priors for  $\beta$ ,  $\gamma$  and  $\kappa$ , and Uniform(0,1) for the ascertainment rate. Four sensitivity analyses were performed: (1) a model with a variable contact rate in three roughly equal periods, i.e. until the lockdown, until 10 February and until 23 February, (2) variable contact rate and an assumed 10% ascertainment rate, (3) assuming a 85% drop in contact rate after the lockdown, as recently suggested [13], and (4) a simplified SIR model (rather than SEIR). All analyses were done with R and JAGS, using 4 chains of 20000 iterations each and another 5000 as burn-in. Model estimates are presented as posterior medians and 95% Credible Intervals (CrI). The full analysis code is available on GitHub: <https://github.com/thlytras/nCoVasc>.

## Results

A total of 49,948 confirmed COVID-19 cases were reported in Wuhan until 9 March, of whom 46,159 were reported until February 23. Figure 2 illustrates the confirmed cases and the estimated epidemic curves by symptom onset and infection date. All model results for the main and sensitivity analyses are detailed on Table 1. The ascertainment rate was estimated at just 0.465% (95% CrI 0.464–0.466%), meaning that for every confirmed case in Wuhan there had been approximately 200 SARS-CoV-2 infections that were not detected. The  $R_0$  was estimated at 3.07 (95% CrI 3.02–3.11), and the incubation period and serial interval at 4.4 and 8.8 days respectively; the latter remained identical in sensitivity analyses.

Assuming a variable contact rate over the study period, the  $R_0$  estimate would be as high as 5.33 until the Wuhan lockdown, falling to 2.09 until 10 February then up again to 3.83, with an ascertainment rate of 0.48%. A nearly identical rate of 0.51% was obtained with a simplified SIR model, and a slightly higher rate of 0.93% assuming an 85% drop in contact rate after the lockdown. On the other hand, if one assumes a fixed ascertainment rate of 10% and a variable contact rate, the  $R_0$  would need to be as low as 0.85 right after the lockdown, and yet lower to 0.45 after 10 February, in order to fit the data (Table 1, sensitivity analysis 2).

The projected course of the epidemic in Wuhan, under the main analysis model, is shown in Figure 3. We estimate that half the population was already infected by 27 January, and by 23 February only 585,000 susceptibles remained (95% CrI 555,000–617,000), meaning 94.7% of the population had been infected. Assuming a 85% drop in contact rate after the lockdown (sensitivity analysis 3), the same percentage drops to 52.7% (95% CrI 51.0–54.4%).

## **Implications for the COVID-19 pandemic**

It has been suggested that the COVID-19 outbreak in Wuhan was controlled via non-pharmaceutical measures and bold public health action that “changed the course of a rapidly escalating epidemic” [11]. Our modelling suggests another plausible scenario: that the large majority of the population was infected, with consequent depletion of susceptibles, and only a very small percentage of infections were ascertained. Otherwise, if ascertainment is assumed higher or not taken into account, very large reductions in transmissibility of SARS-CoV-2 are needed to account for the observed data, with about 80% of infections prevented immediately after the lockdown and 90% after February 10. It should be emphasized that in no way do our results diminish the value of non-pharmaceutical public health interventions in mitigating the outbreak, nor do they underestimate the immense efforts of the Chinese people. Their bold, large-scale action undoubtedly alleviated an already very difficult situation, and likely succeeded in “flattening the curve” of the pandemic to the maximum possible extent.

Accounting for case ascertainment is crucial for accurately assessing the true nature of the COVID-19 pandemic. An ascertainment rate as low as 0.5% in Wuhan suggests that the IFR of the SARS-CoV-2 virus is orders of magnitude lower than the CFR of COVID-19 (although some underascertainment of COVID-19 deaths should be taken into account as well). Together with a  $R_0$  estimate of 3.07, it would make COVID-19

the archetype of a high-transmissibility / low-severity pandemic, whose high impact is driven by the sheer number of infections rather than the virulence of the pathogen [14]. Our result further implies a very large amount of transmission from asymptomatic or minimally symptomatic persons, raising doubts about the feasibility of suppressing, much less containing, the COVID-19 pandemic [15]. Currently the World Health Organization (WHO) still advocates for a “containment strategy”, with laboratory testing for every suspected case, isolation of confirmed cases (including mild cases) in health facilities, and tracing all contacts [16]. Laboratory testing is indeed a priority for surveillance purposes, and also for severely ill patients, healthcare workers and other selected groups [17]. Given our findings, however, exhaustive testing and contact tracing are probably not optimal at this point in countries with ongoing epidemics, and risk detracting from the only eminently achievable end goal, which is mitigation: “flattening the curve” as much as possible, reducing stress on healthcare facilities and protecting the vulnerable. That is what is likely occurring in other locked-down areas of China right now, with the virus on a slow and stealthy progression, rather than on a freeze. Given the severity of the pandemic threat, no effort should be spared and no public health intervention discounted a priori. But at the same time, it is crucial to set realistic priorities and realistic expectations, as well as allocate resources optimally while advocating for more.

Our estimated ascertainment rate is on the lower end of previously published estimates for Wuhan [4]. Regardless of the precise rate though, the main conclusion about the feasibility of “containment/suppression” (or lack thereof) still holds, as also evidenced by the rapid worldwide spread of SARS-CoV-2. Our result also highlights the urgent need for seroprevalence surveys in different populations, to formally elucidate the true prevalence of the infection and guide further management of this new global pandemic. Striving for more timely data remains essential for optimal public health decisions.

## **Acknowledgements**

We gratefully acknowledge Dr. Vana Sypsa for providing very useful comments on this manuscript

## **Conflicts of Interest**

None

## **Funding**

No funding was received for this study

## References

1. Kobayashi T, Jung S, Linton NM, Kinoshita R, Hayashi K, Miyama T, et al. Communicating the Risk of Death from Novel Coronavirus Disease (COVID-19). *Journal of Clinical Medicine*. 2020 Feb;9(2):580.
2. Read JM, Bridgen JR, Cummings DA, Ho A, Jewell CP. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. medRxiv. 2020 Jan 28;2020.01.23.20018549.
3. Nishiura H, Kobayashi T, Yang Y, Hayashi K, Miyama T, Kinoshita R, et al. The Rate of Underascertainment of Novel Coronavirus (2019-nCoV) Infection: Estimation Using Japanese Passengers Data on Evacuation Flights. *J Clin Med*. 2020 Feb 4;9(2).
4. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med [Internet]*. 2020 Mar 19 [cited 2020 Mar 23]; Available from: <http://www.nature.com/articles/s41591-020-0822-7>
5. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*. 2020 Mar 16;
6. Tuite A, Ng V, Rees E, Fisman D. Estimation of COVID-19 outbreak size in Italy based on international case exportations. medRxiv. 2020 Mar 6;2020.03.02.20030049.
7. Tuite AR, Bogoch I, Sherbo R, Watts A, Fisman DN, Khan K. Estimation of COVID-2019 burden and potential for international dissemination of infection from Iran. medRxiv. 2020 Feb 25;2020.02.24.20027375.
8. Wu T, Ge X, Yu G, Hu E. Open-source analytics tools for studying the COVID-19 coronavirus outbreak. medRxiv. 2020 Mar 5;2020.02.25.20027433.
9. Backer JA, Klinkenberg D, Wallinga J. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20-28 January 2020. *Euro Surveill*. 2020 Feb;25(5).
10. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*. 2020 Jan 29;
11. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19) [Internet]. 2020 Feb [cited 2020 Mar 5] p. 40. Available from: <http://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>
12. Daley DJ, Gani JM. *Epidemic modelling: an introduction*. Cambridge ; New York: Cambridge University Press; 1999. 213 p. (Cambridge studies in mathematical biology).
13. Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Age profile of susceptibility, mixing, and social distancing shape the dynamics of the novel coronavirus disease 2019 outbreak in China. medRxiv. 2020 Mar 20;2020.03.19.20039107.
14. Reed C, Biggerstaff M, Finelli L, Koonin LM, Beauvais D, Uzicanin A, et al. Novel framework for assessing epidemiologic effects of influenza epidemics and pandemics. *Emerging Infect Dis*. 2013 Jan;19(1):85–91.
15. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*. 2020 Mar 16;
16. World Health Organization. WHO Emergencies Press Conference on coronavirus disease outbreak - 16 March 2020 [Internet]. Available from: <https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencies-coronavirus-press-conference-full-16mar2020.pdf>

17. European Centre for Disease Prevention and Control. Rapid risk assessment: Novel coronavirus disease 2019 (COVID-19) pandemic: increased transmission in the EU/EEA and the UK – sixth update [Internet]. European Centre for Disease Prevention and Control. 2020 [cited 2020 Mar 18]. Available from: <https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-novel-coronavirus-disease-2019-covid-19-pandemic-increased>

**Table 1: Model-based estimates for the COVID-19 epidemic parameters in Wuhan, China.**

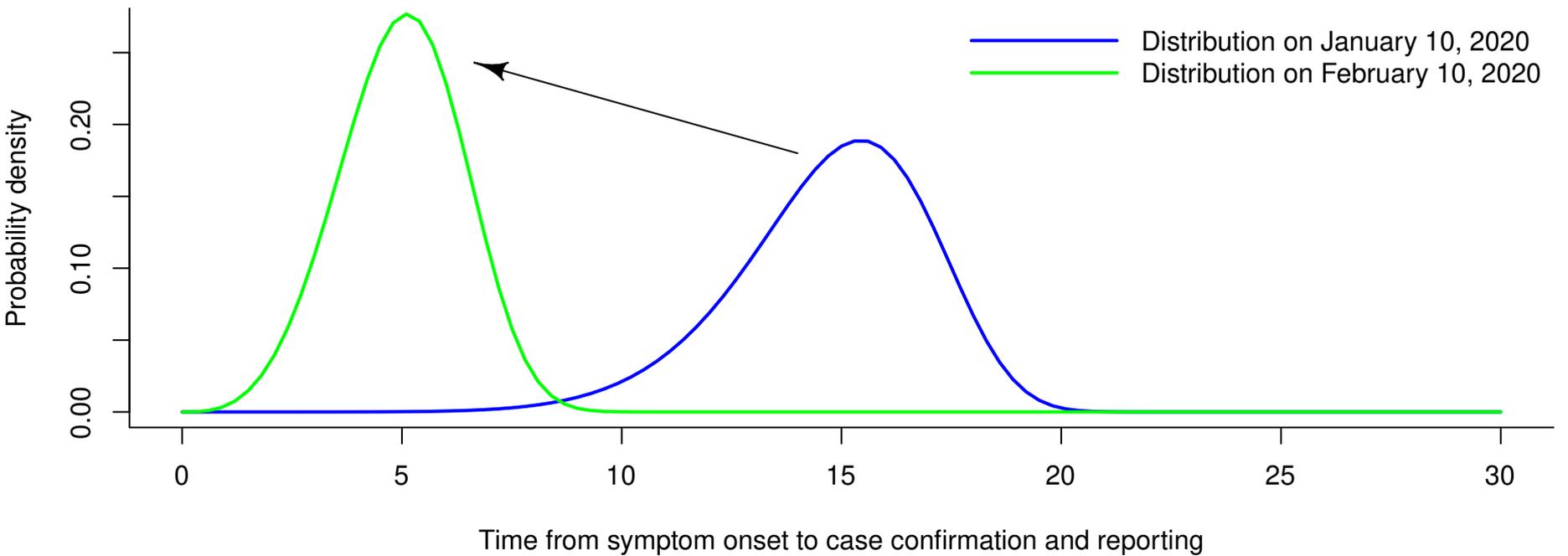
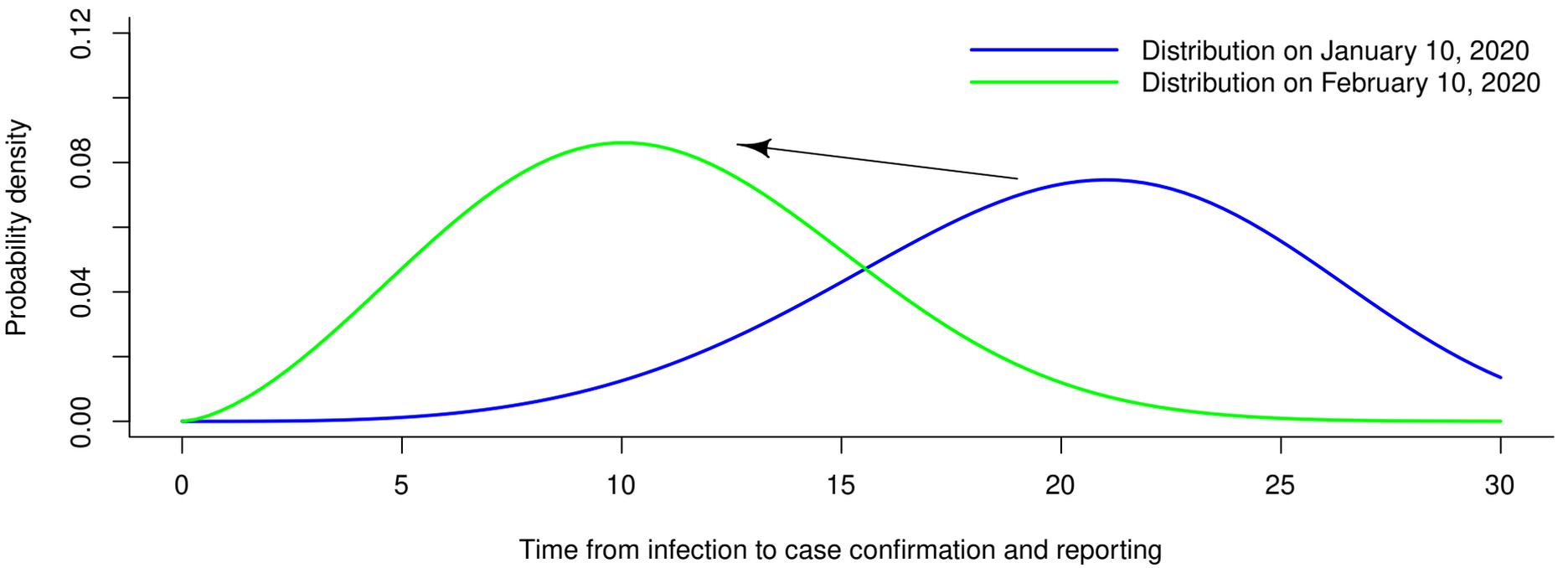
Parameter	Estimate (95% CrI)
<i>Main analysis</i>	
Ascertainment rate (%)	0.465 (0.464–0.466)
Basic Reproduction Number ( $R_0$ )	3.07 (3.02–3.11)
Incubation period (days)	4.38 (4.34–4.41)
Serial interval (days)	8.80 (8.75–8.87)
<i>Sensitivity analysis 1 (variable contact rate)</i>	
Ascertainment rate (%)	0.480 (0.477–0.484)
Basic Reproduction Number ( $R_0$ ) (until lockdown)	5.33 (5.23–5.42)
Basic Reproduction Number ( $R_0$ ) (until Feb 10)	2.09 (2.05–2.13)
Basic Reproduction Number ( $R_0$ ) (until Feb 23)	3.83 (3.62–4.04)
Incubation period (days)	4.38 (4.34–4.42)
Serial interval (days)	8.81 (8.75–8.87)
<i>Sensitivity analysis 2 (variable contact rate, and 10% ascertainment rate assumed)</i>	
Basic Reproduction Number ( $R_0$ ) (until lockdown)	4.13 (4.06–4.20)
Basic Reproduction Number ( $R_0$ ) (until Feb 10)	0.85 (0.84–0.87)
Basic Reproduction Number ( $R_0$ ) (until Feb 23)	0.45 (0.44–0.46)
Incubation period (days)	4.38 (4.34–4.41)
Serial interval (days)	8.81 (8.75–8.87)
<i>Sensitivity analysis 3 (assuming 85% drop in contact rate after lockdown)</i>	
Ascertainment rate (%)	0.930 (0.899–0.965)
Basic Reproduction Number ( $R_0$ ) (until lockdown)	5.86 (5.76–5.97)
Basic Reproduction Number ( $R_0$ ) (until Feb 10)	0.88 (0.86–0.89)
Incubation period (days)	4.38 (4.33–4.41)
Serial interval (days)	8.81 (8.75–8.87)
<i>Sensitivity analysis 4 (simplified SIR model)</i>	
Ascertainment rate (%)	0.513 (0.510–0.516)
Basic Reproduction Number ( $R_0$ )	2.29 (2.25–2.32)
Serial interval (days)	13.87 (13.74–14.00)

## Figure Legends

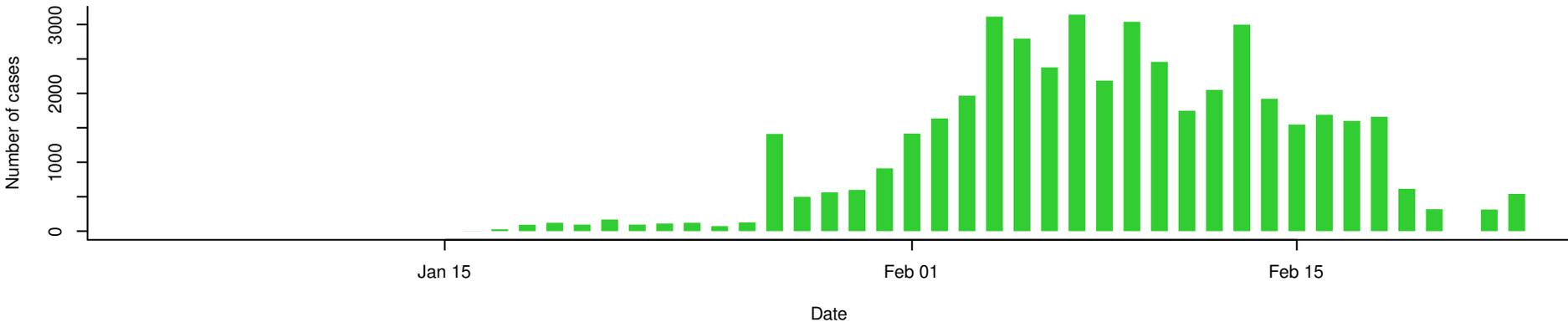
Figure 1: Assumed distributions for time between case infection and case reporting, and time between symptom onset and case reporting

Figure 2: Cases of COVID-19 in Wuhan by reporting date, symptom onset (estimated) and time of infection (estimated)

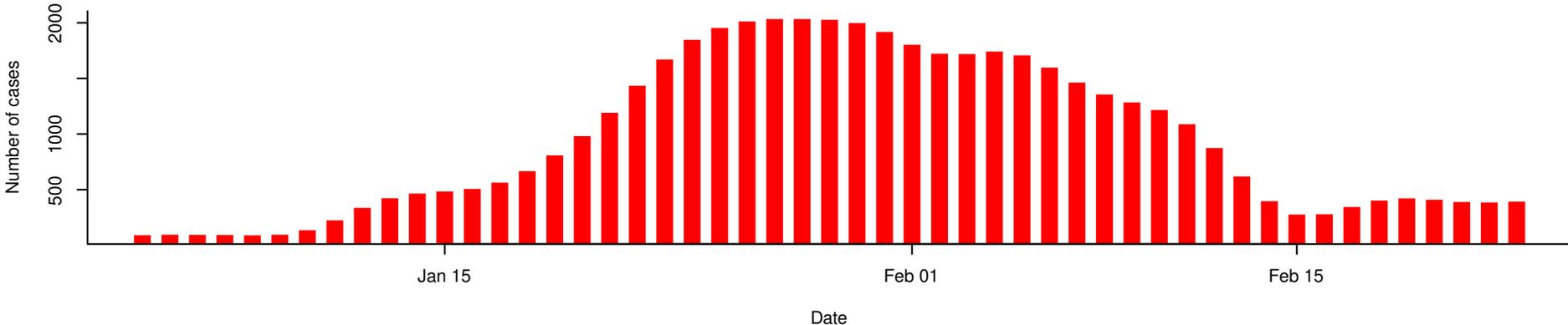
Figure 3: Estimated Susceptible, Exposed, Infected and Recovered population from COVID-19 over time in Wuhan



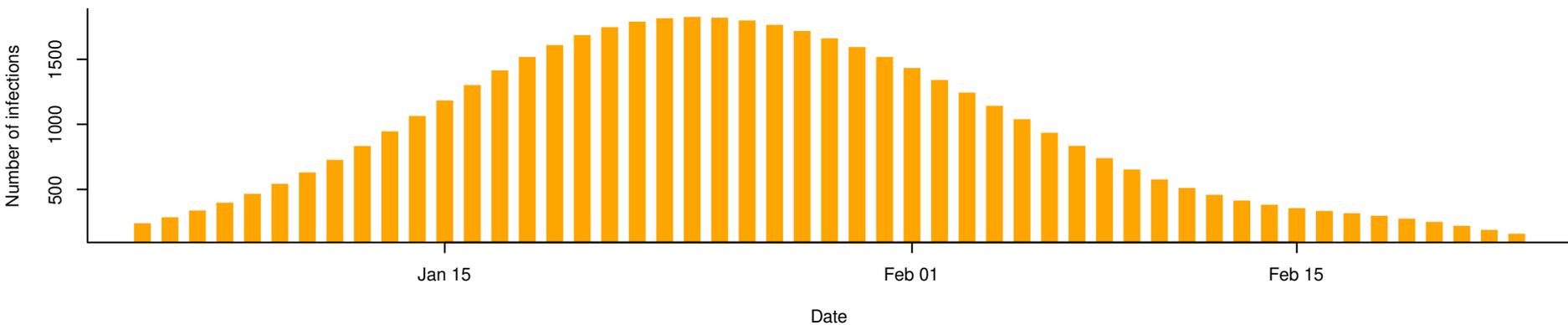
### Reported cases in Wuhan



### Estimated symptomatic cases in Wuhan



### Estimated infections in Wuhan



# Estimated course of the COVID-19 epidemic in Wuhan

