Global underreporting of COVID-19 cases during 1 January 2020 to 6 May 2022

Globally, the number of reported COVID-19 cases from December 2019 to 6 May 2022 was 513,955,910 and the reported number of deaths during this period was 6,190,349 (ref. 1). The number of hospitalized cases all over the world due to the pandemic started declining at the beginning of April 2022 (ref. 2). Globally, the number of fully vaccinated individuals as of 6 May 2022 was 5.1 billion (ref. 1). Vaccinated individuals have higher chances of not being hospitalized[2], but they still have some probability of acquiring the virus[3]. This indicates that COVID-19 transmission occurs between vaccinated and unvaccinated individuals.

The present study aims to provide a range of model-based estimates of the total number of COVID-19 cases in the world from the beginning of the pandemic to 6 May 2022 and the proportion of the reported cases. The methodology used is similar to the model-based estimation of underreporting of COVID-19 cases in eight countries through 31 March 2020 (ref. 4). The degree of underreporting could have varied over the period 2020–22. There could be one or more reasons for this, namely (i) all COVID-19 positive cases have not been tested, and (ii) not all those who tested positive for COVID-19 have been reported. In the past, there have been methods to understand the extent of underreporting in various epidemics. However, the supportive data collection for such purposes is limited[5–8].

The estimated world population in 2022 is 7.954 billion (ref. 8). This is divided into 1.989 billion (aged 0–14 years), 5.170 billion (aged 15–65 years) and 795 million (more than 65 years)[9]. The estimated world population in 2020 was 7.794 billion. The average population for 2020–22 is 7.874 billion and the average annual exponential growth rate of the population during this period is $(1/2)\log(P_2/P_0) = 0.0044$. Here P_0 and P_2 denote the population in 2020 and 2022 respectively. The proportion of average COVID-19 infections against the effective population during 2020–22 is approximately 0.065. This value (0.065) is obtained as the proportion of the number of reported COVID-19 cases as of 6 May 2022, viz. 513, 955, 910 to the average global population for 2020–22, which is 7.874 billion. This indicates that on an average, each individual in the world has contributed to infections at the rate of 0.065 during 2020–22. This rough estimate of the global average force of infection suggests that every 1000 individuals in the world contributed 65 new infections. This average estimate will vary on several factors, such as demographic characteristics (e.g. age, sex, race), socio-economic status (e.g. occupation, income and education), public health measures (e.g. mask-wearing, closures and lockdowns) and medical conditions (e.g. congestive heart failure and diabetes). The figure 65/1000 only indicates the global reported number, which is underreported.

We have generated model-based COVID-19 infection data as of 6 May 2022 (Table 1). We have adopted the methodology in our previous study on underreporting[4]. The transmission parameter β was initially considered 2.0×10^{11} and simulations were performed for 121 weeks from 1 January 2020. The transmission parameter indicates the average number of susceptible people that one infected person will infect during a unit period. We have provided ten additional model-based simulations performed for $\beta \pm \beta \cdot k\%$ for $k = 1, 2, \ldots, 5$ (Table 1). The Appendix provides technical details.

With an approximate population of 5.1 billion fully vaccinated in the world, the spread of COVID-19 might have been reduced. Yet the debate on the actual magnitude and number of infected individuals during 2020–22 is still not over. We estimate that the number of COVID-19 infections from December 2019 to 6 May 2022 is between 601 million and 2.4 billion. This wide range is indicative of the factors that have affected the reporting of COVID-19 cases across different countries, which include deliberate data-tampering, both malicious and otherwise, the inability to conduct accurate case tracking and the lack of uniformity in reporting, among others[8].

Our goal here is to provide model-based estimates and predictions of global underreporting of COVID-19 cases during a critical period of the pandemic. There is evidence that the widely advertised global case counts are quite inaccurate because of the lack of information. We have explored and explained some of these phenomena.

Appendix: Transmission dynamics and wavelets.

We have adopted standard susceptible $S(t)$ and infected $I(t)$ dynamics through direct interaction and with the average transmission rate parameter β. Since COVID-19 was seen as a global pandemic without exception, we have considered $S(t) + I(t) = N$. The populations are coupled with two differential equations:

$$dS(t)/dt = -\beta S(t)I(t) \text{ and}$$

$$dI(t)/dt = \beta S(t)I(t).$$

The initial value of susceptible is the same as the total population at the beginning of 2020 described in the text. As of 1 January 2020 assumed there were 10 infected individuals globally. We do not have any authentic data on the actual infected individuals as of 1 January 2020. Most of our methods are similar to the ones we have used in an earlier study[4]. The Meyer wavelets are flexible in nature. The level of reporting of

Table 1. Model-based COVID-19 global estimates as of 6 May 2022 and the level of underreporting

COVID-19 estimates	$\beta \pm \beta \cdot k$	Level of reporting
2,429,163,263	$2.0 \times 10^{11} + 2.0 \times 10^{11} \times 0.05$	1 in 4.7
2,126,135,922	$2.0 \times 10^{11} + 2.0 \times 10^{11} \times 0.04$	1 in 4.1
1,848,079,423	$2.0 \times 10^{11} + 2.0 \times 10^{11} \times 0.03$	1 in 3.6
1,596,288,176	$2.0 \times 10^{11} + 2.0 \times 10^{11} \times 0.02$	1 in 3.1
1,370,997,166	$2.0 \times 10^{11} + 2.0 \times 10^{11} \times 0.01$	1 in 2.7
1,171,568,510	2.0×10^{11}	1 in 2.3
996,703,028	$2.0 \times 10^{11} - 2.0 \times 10^{11} \times 0.01$	1 in 1.9
844,648,392	$2.0 \times 10^{11} - 2.0 \times 10^{11} \times 0.02$	1 in 1.6
713,384,605	$2.0 \times 10^{11} - 2.0 \times 10^{11} \times 0.03$	1 in 1.4
600,776,595	$2.0 \times 10^{11} - 2.0 \times 10^{11} \times 0.04$	1 in 1.2
504,690,975	$2.0 \times 10^{11} - 2.0 \times 10^{11} \times 0.05$	1 in 1

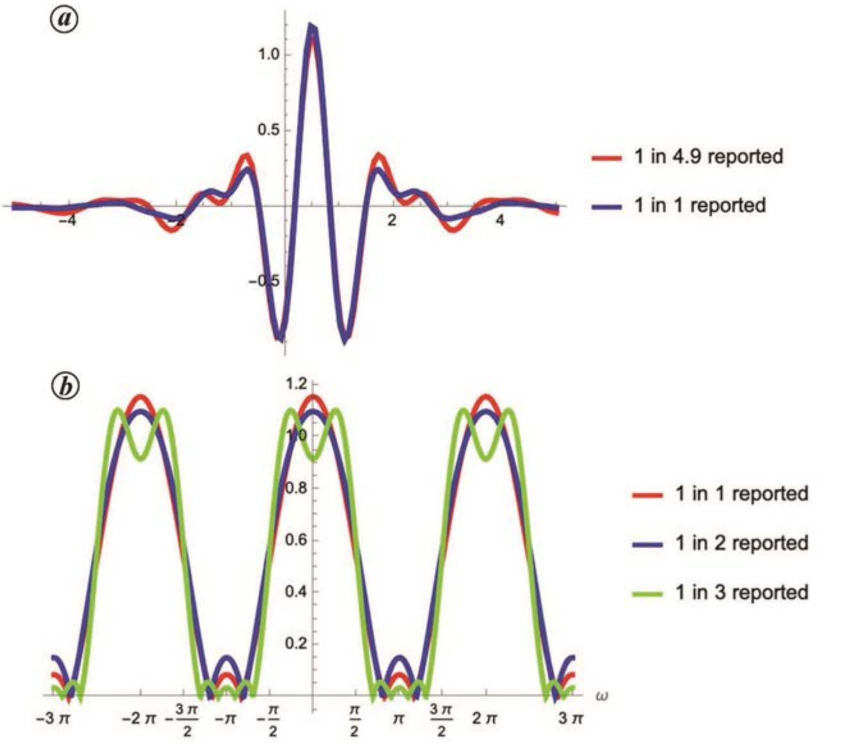

Figure 1. ***a***, Meyer wavelets of different levels of underreporting. ***b***, Frequency response plots for Meyer wavelets of different levels of underreporting.

COVID-19 cannot be uniform over the two years for the reasons explained in the main text. Here we have computed the level of reporting as the ratio of model-based predictions (as of 6 May 2022) to the global cumulative reported numbers for that day. The Meyer wavelets $\psi(\omega)$ are infinitely differentiable functions within a certain domain[4]. They are accompanied by a function $\nu(x)$ given by

$$\psi(\omega)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}\sin\left(\dfrac{\pi}{2}\nu\left(\dfrac{3|\omega|}{2\pi}-1\right)\right)e^{\frac{i\omega}{2}}, \\ \quad \text{if } \; 2\pi/3<|\omega|<4\pi/3 \\ \dfrac{1}{\sqrt{2\pi}}\cos\left(\dfrac{\pi}{2}\nu\left(\dfrac{3|\omega|}{2\pi}-1\right)\right)e^{\frac{i\omega}{2}}, \\ \quad \text{if } \; 4\pi/3<|\omega|<8\pi/3 \\ \quad 0 \quad \text{otherwise}\end{cases}$$

Here $\nu(x) = 0$ for $x < 0$, and

$$\nu(x)=\begin{cases}x & \text{for } x\in(0,1)\\ 1 & \text{for } x>1.\end{cases}$$

The frequency response for $\psi(t)$ is given by

$$\sum_k b_k e^{-ik\omega}.$$

We have plotted Meyer wavelets for different levels of reporting for the range $\omega \in [-3\pi, 3\pi]$.

The proposed method blends susceptible and infected model (SI model) of disease spread with Meyer wavelets. The SI model has only one parameter called 'transmission parameter' with two variables, namely susceptible and infected. The SI model equations can also be written as the following two ordinary differential equations

$$\frac{dS(t)}{dt}=\frac{-\beta S(t)I(t)}{N},$$

$$\frac{dI(t)}{dt}=\frac{\beta S(t)I(t)}{N}=\beta I(t)\left(1-\frac{I(t)}{N}\right),$$

where $S(t) + I(t) = N$. Due to the presence of the term $(1 - (I(t)/N))$ this model resembles the logistic growth model.

Once model-based estimates are obtained, the Meyer wavelets are used to demonstrate the underreporting of COVID-19 cases in the population (Figure 1).

1. WHO Coronavirus (COVID-19) Dashboard, WHO Coronavirus (COVID-19) Dashboard with Vaccination Data; https://covid19.who.int/ (accessed on 6 May 2022).
2. Coronavirus (COVID-19) Hospitalizations – Our World in Data; https://ourworldindata.org/covid-hospitalizations (accessed on 6 May 2022).
3. Plumb, I. D. *et al.*, *MMWR Morb. Mortal. Wkly. Rep.*, 2022, **71**, 549–555; doi:http://dx.doi.org/10.15585/mmwr.mm7115e2external icon.
4. Krantz, S. and Rao, A., *Infect. Control Hosp. Epidemiol.*, 2020, **41**(7), 857–859; doi:10.1017/ice.2020.116.
5. Krantz, S. G., Polyakov, P., Rao, A. S. R. S., *J. Theor. Biol.*, 2020, **494**, 110243.
6. Angulo, F. J., Finelli, L. and Swerdlow, D. L., *JAMA Netw. Open*, 2021, **4**(1), e2033706; doi:10.1001/jamanetworkopen.2020.33706.
7. Srinivasa Rao, A. S. R., *J. Theor. Biol.*, 2012, **302**, 89–95.
8. Kobak, D., *Significance*, 2022, **19**(2), 10–13.
9. UN World Population Dashboard, United Nations Population Fund (https://www.unfpa.org/data/world-population-dashboard).

ACKNOWLEDGEMENTS. We thank the anonymous reviewer for critical comments that helped improve the manuscript. S.G.K. and A.S.R.S.R. were supported by NSF grant DMS-2140493.

Received 9 May 2022; revised accepted 3 August 2022

STEVEN G. KRANTZ[1]
DAVID A. SWANSON[2]
ARNI S. R. SRINIVASA RAO[3,*]

[1]*Washington University in St. Louis, Campus Box 1146, One Brookings Drive, St. Louis, Missouri 63130, USA*
[2]*Department of Sociology, University of California, Riverside, CA 92521, USA and Center for Studies in Demography and Ecology, University of Washington, Seattle, WA 98195, USA*
[3]*Laboratory for Theory and Mathematical Modeling, Department of Medicine–Division of Infectious Diseases, Medical College of Georgia, 1120 15th Street, AE2040, Augusta, GA 30912, USA*
**For correspondence.*
e-mail: arrao@augusta.edu