

Article Type: Research Brief

Words: 873 (without title, authors, affiliations, references, Table & Figure)

Level of under-reporting including under-diagnosis before the first peak of COVID-19 in various countries: Preliminary Retrospective Results Based on Wavelets and Deterministic Modeling

Steven G. Krantz

Professor,

Department of Mathematics,

Washington University in St. Louis, Campus Box 1146,

One Brookings Drive, St. Louis, Missouri 63130, USA.

Arni S.R. Srinivasa Rao

Professor, Division of Health Economics and Modeling,

Department of Population Health Sciences,

Medical College of Georgia, Augusta University, Augusta, USA.

Director, Laboratory for Theory and Mathematical Modeling,

Division of Infectious Diseases, Department of Medicine,

Medical College of Georgia, Augusta, GA 30912, USA.

Department of Mathematics, Augusta University, Augusta, GA 30912, USA.

*Corresponding author (email: arrao@augusta.edu)

We have estimated under-reporting of the novel coronavirus or COVID-19 until March 9, 2020, in various countries until the first peak occurred in each of the countries that have reported 500 or more cases of COVID-19 as of March 9, 2020. Our retrospective model-based estimations of under-reporting (including those due to under-diagnosis) will be helpful in the assessment of pandemic preparedness required. The ratio of reported COVID-19 cases to model-based predictions of COVID-19 for eight major countries who have reported 500 or more cases up to March 9, 2020, were provided (See Column I in Table 1). COVID-19 Reporting in France, Germany Italy, and South Korea was comparatively much better than other countries. For the U.S. The data until March 9, 2020 was not sufficient to provide a robust estimate.

There have been 109,000 cases of COVID-19 and 3800 related deaths across the world according to the situational report-49 of the World Health Organization (WHO) released on March 9, 2020¹. The majority of these cases---around 80,700---were from China and eight more countries, namely, Italy, South Korea, Iran, France, Germany, Spain, the USA, and Japan. These countries have reported 500 or more confirmed cases of COVID-19¹⁻². Identification of possible cases of COVID-19 is arguably more important in controlling high traffic to hospitals and emergency departments³. Earlier models on COVID-19 did reflect the importance of data collection⁴.

Actual pandemic preparedness depends on true cases in the population, whether or not they are identified. Preventing transmission to the susceptible from these true cases depends on how well we can understand under-reported and under-diagnosed situations promptly. Retrospective analysis of the data will be useful for the next

epidemic but not for the current epidemic. Hence we are proposing to use our methods, which we have been developing in recent years, to provide model-based estimates of under-reporting for COVID-19 within a few weeks.

New methods using harmonic analysis and wavelets which we are developing---some of them just accepted---will be of timely use⁵. We are proposing a model-based evaluation of under-reporting of coronavirus (COVID-19) in various countries across the world using the methods which we have recently developed with harmonic analysis⁵, i.e. develop full epidemic data from partial data (using wavelet approach). However, the current article is a preliminary analysis and modeling done on the data so far available up to March 9, 2020. As of the article preparation day, it may not represent the pandemic in its entire scale until it is completely controlled. The predictions provided on under-reporting until March 9 in a couple of European countries were found to be close to reality as more COVID-19 cases surfaced from March 9 to March 16, 2020. Wavelets of reported and adjusted to the under-reporting cases are shown in Figure 1. We are also anticipating, in the long run, to use other techniques⁵⁻⁹ for understanding the reporting once more data is available.

Data, Methods & Models

We have collected COVID-19 and population data in each country for our analysis from the WHO¹, Worldometer² and World Bank¹⁰ sources. We have used population densities, proportion of the population living in urban areas, and populations delineated by three age groups, namely, 0-14, 15-64, 65+. Further, we considered daily new cases (above 10) up to seeing the first reported peak of COVID-19 cases and the corresponding date ranges for all the countries for which such data available. This range of days varied between 8 and 16 days (Table 1). We use two coupled differential equations $\dot{s}(t) = -\beta s(t)k(t)$ and $\dot{k}(t) = \beta s(t)k(t)$, where $s(t)$ and $k(t)$ represent susceptible and infected at time t , and β is the transmission rate

that is assumed to be invariant within the range of days for which the infection numbers in each country were computed. The β values for the age groups 15-64 and 65+ considered for various countries are as follows: China: $\{(0.8 \times 1.50E-08, 1.50E-08), 7.5E-09\}$, Italy: $\{1.50E-08, 3.00E-08\}$, Iran: $\{1.50E-08, 9.00E-08\}$, South Korea: $\{2.25E-08, 4.50E-08\}$, France: $\{1.50E-08, 3.00E-08\}$, Spain: $\{3E-08, 6.00E-08\}$, Germany: $\{1.50E-08, 3.00E-08\}$, and the U.S.A.: $\{7.5E-09, 1.50E-08\}$. The difference between model-predicted numbers and the actual numbers reported within the range were treated as under-reported (that includes under-diagnosed). Further, we have constructed the Meyer wavelets for the reported and adjusted data after adjusting the infected number in the population for under-reporting. The Meyer wavelet is a differentiable function, $\psi(\omega)$, which is infinitely differentiable in the domain with a function u given below

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

Here $u(x) = 0$ for $x < 0$, $u(x) = x$ for $x \in (0,1)$, and $u(x) = 1$ for $x > 1$. For further details, please refer to^{5,9}.

As of March 16, 2020, we do not have much information on COVID-19 on its transmissibility rates from infected to uninfected based on migration of populations to construct country wide networks. We also have no clear idea on time to remain active by COVID-19 in non-living objects, etc. So we did not consider the interaction between humans to the non-living surfaces--- such as plastics, metals, papers, etc. Mathematical modeling can be easily made more complex by adding more parameters but caution need to be taken for obtaining parameters with scientifically collected data and readily available parameters from well-designed studies. Once more data on networks, duration of living of COVID-19 on non-living objects, we can build more complex models with more parameters.

References:

1. [The WHO Situational Report-49](#) (accessed on March 9, 2020)
2. Coronavirus, Worldometer, <https://www.worldometers.info/coronavirus/#countries> (accessed on March 11, 2020)
3. Rao, A.S.R.S, & Vazquez, J.A. (2020). Identification of COVID-19 Can be Quicker through Artificial Intelligence framework using a Mobile Phone-Based Survey in the Populations when Cities/Towns Are Under Quarantine. *Infection Control & Hospital Epidemiology*, (Cambridge) 1-18. doi:10.1017/ice.2020.61
4. Wu, JT, Leung, K, Leung, GM (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, *The Lancet*, Volume 395, Issue 10225,2020,Pages 689-697,ISSN 0140-6736,[https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).
5. Krantz S.G., Polyakov P.G. Rao, A.S.R.S. (2020). True epidemic growth construction through harmonic analysis. *Journal of Theoretical Biology* (Elsevier) [Track paper](#).
6. Rao, A.S.R.S. (2012). Understanding theoretically the impact of reporting of disease cases in epidemiology. *Journal of Theoretical Biology* (Elsevier) 302 (2012), 89–95.
7. Atkins Katherine E, Wenzel Natasha S, Ndeffo-Mbah Martial, Altice Frederick L, Townsend Jeffrey P, Galvani Alison P et al. Under-reporting and case fatality estimates for emerging epidemics *BMJ* 2015; 350 :h1115
8. Gamado, K.M., Streftaris, G. & Zachary, S. Modelling under-reporting in epidemics. *J. Math. Biol.* 69, 737–765 (2014). <https://doi.org/10.1007/s00285-013-0717-z>
9. Krantz, S.G. (1999). *A Panorama of Harmonic Analysis*, The Carus Mathematical Monographs, Number 27, Mathematical Association of America.
10. The World Bank Open Data. <https://data.worldbank.org/> (accessed on March 11, 2020)

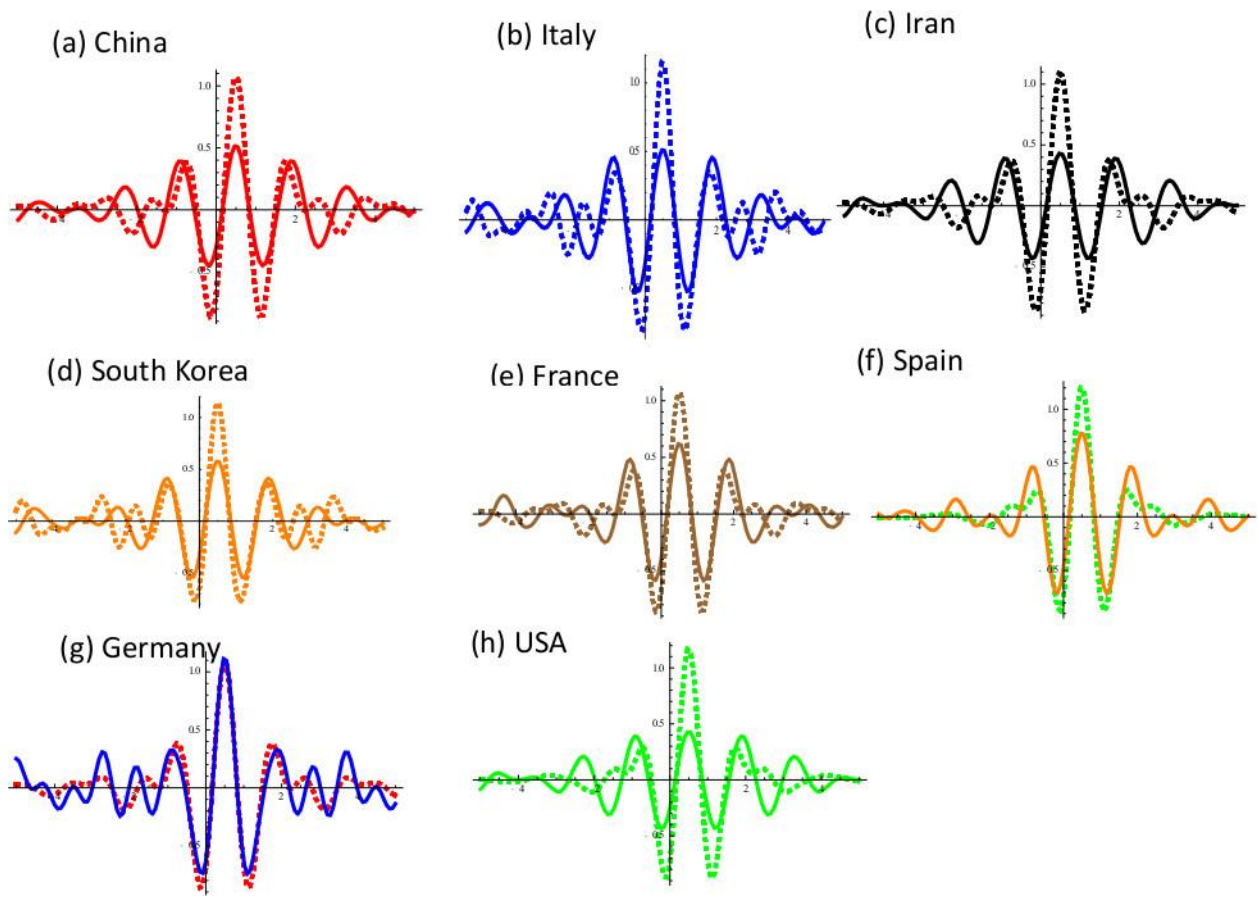
Acknowledgements: We thank Editor-in-Chief, Associate Editor (Handling) and the Statistical Consultant for their constructive comments which helped us in thoroughly revising our paper.

Authors contributions: Both the authors contributed in writing. ASRS Rao designed the study, developed the methods, collected data, performed analysis, computing, wrote the first draft. SG Krantz designed the study, contributed in writing, performed analysis, editing the draft.

Conflicts: None.

Funding: None to report to this study.

Figure 1. Meyer wavelets for various countries for reported (dashed lines) and adjusted data after adjusting for under-reporting listed in the Table 1.



Country (in the order of confirmed cases)	Total COVID- 19 Cases	Total deaths	Total Population Density (P/Km ²) Year 2020	Urban Population (%) Year 2020	Dates range of daily new cases up to the first peak of such cases	Daily New cases range up to the first peak	Proporti on of Populati on in the 0-14 age group (year 2018) (%)	Proporti on of Populati on in the 15-65 age group (year 2018) (%)	Proporti on of Populati on aged 65 and above in 2018 (%)	Model- based Under reported /diagnos ed Up to March 9, 2020	Number of people reported To the number infected
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
China	80,761	3,136	153	61	Jan 22 – Feb 4	259- 3884	17.9	71.2	10.9	12.03 – 89.2 millions	1 in 149 to 1 in 1104
Italy	10,149	631	206	69	Feb 22 – Mar 9	58- 1797	13.3	64.0	22.7	30223	1 in 4 reported
Iran	8,042	291	52	76	Feb 21 – Mar 6	13 – 1234	24.5	69.3	6.2	266213	1 in 34 reported
South Korea	7,513	58	527	82	Feb 23 – Mar 3	27 – 851	13.0	72.6	14.4	18809	1 in 4 reported
France	1,784	33	119	82	Feb 27 – Mar 7	20 – 296	18.0	62.0	20.0	7931	1 in 5 reported
Spain	1,690	35	94	80	Feb 27 – Mar 9	12 – 557	14.7	66.0	19.3	87405	1 in 53 reported

Country (in the order of confirmed cases)	Total COVID- 19 Cases	Total deaths	Total Population Density (P/Km ²) Year 2020	Urban Population (%) Year 2020	Dates range of daily new cases up to the first peak of such cases	Daily New cases range up to the first peak	Proporti on of Populati on in the 0-14 age group (year 2018) (%)	Proporti on of Populati on in the 15-65 age group (year 2018) (%)	Proporti on of Populati on 65 and above in 2018 (%)	Model- based Under reported /diagnos ed Up to March 9, 2020	Number of people reported To the number infected
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
Germany	1,458	2	240	76	Feb 27 – Mar 5	22 – 283	13.6	65.0	21.4	2277	1 in 3 reported
USA	874	28	36	83	Mar 2 – Mar 12	25 – 1652	18.7	65.5	15.8	1.21 million (not sufficient data)	1 in 406 reported (not sufficient data)

Table 2. COVID-19 cases, demographics, daily cases, growth rates, estimated under-reporting up to March 9, 2020.