ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - 2023, Volume 18, Issue 1

What is wrong with Chinese COVID-19 statistics?

Igor Nesteruk(1) (2)

(1) Institute of Hydromechanics, National Academy of Sciences of Ukraine, Kyiv, Ukraine.
(2) Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine.

CORRESPONDING AUTHOR: Igor Nesteruk, nesteruk@yahoo.com, ORCID 0000-0001-7250-2729.

SUMMARY

The media is reporting tens of millions of new daily Covid-19 cases in China in the final days of 2022. However, official statistics have recorded in mainland China only1.9 million cases since the start of the pandemic and stopped providing data after December 21, 2022. Results of SIR simulations showed that daily numbers of new cases stated to decline in December 2022. The contradictions in statistics and estimations are discussed. Millions of new daily cases in China look very unlikely.

Keywords: COVID-19 pandemic; epidemic waves; epidemic dynamics in China; mathematical modeling of infection diseases; the generalized SIR model; parameter identification; statistical methods.

According to the information from an internal meeting of China's National Health Commission (NHC) held on December 21, 2022, the 248 million people were infected with Covid-19 in the first 20 days of December with nearly 37 million new cases on a single day [1]. These figures contradict with official statistics, reflected in COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU), [2]. According to the version of the JHU file, available on January 4, 2023, as of December 21, 2022 the accumulated numbers of cases were:1,909,905 in mainland China; 2,402,238 in Hong Kong; 8,624,680 in Taiwan;1,569 in Macao. Taking the sum of this figures, we have the value 12,938,392 that is 19 times lower than NHC estimation of the number of cases accumulated in the first 20 days of December 2022 and almost 3 times lower than its daily maximum figure.These huge discrepancies may indicate either completely incorrect NHC statistical data (which stopped be ingupdated after December 21, 2022, see [2] and Table1), or a new estimate of the real number of cases (the details of which we do not know).

The difference between the real and detected numbers of COVID-19 cases can be huge, [3,4], especially in countries with low testing level [4]. In mainland China the number of accumulated tests per capita TC was 6.46 already on April 11, 2022, [2]. The testing levels in Hong Kong are similar (TC=6.59 as of May 24, 2022, [2]), and approach to the highest values in other countries [2]. Thus, there is no reason to think that many cases in China was not detected even after November 30, when the Zero-Covid [5] restrictions began to loosen [6].

The epidemic dynamics in mainland China in December 2022 can be compared with the results of application of the generalized SIR model, which links the number of susceptible (and unprotected) people S, infectious (infected and spreading the infection) I and removed R (immunized, isolated and dead) over time t, [7,8]. Then the sum $V(t)=I(t)+R(t)$ is the theoretical estimation of the accumulated numbers of cases V_j and is derivative dV/dt is the estimation of the daily numbers of new cases.The values V_j corresponding to the period November 15–28, 2022 were taken in [9] for calculations the optimal SIR parameters according to the method presented in [8]. Since this period was before the beginning of easing the Zero-Covid restrictions in mainland China (November 30, 2022, [6]), the obtained forecast allows us to follow what the epidemic dynamics could be without the easing of the restrictions. Corresponding SIR curves are presented in Fig.1.

DOI: 10.54103/2282-0930/20637
Accepted: 31th January 2023

Fig. 1.The results of the SIR simulations of the CoVID-19 pandemic wave in mainland China in late 2022, [9] and comparisons with the registered values and other calculations (markers).

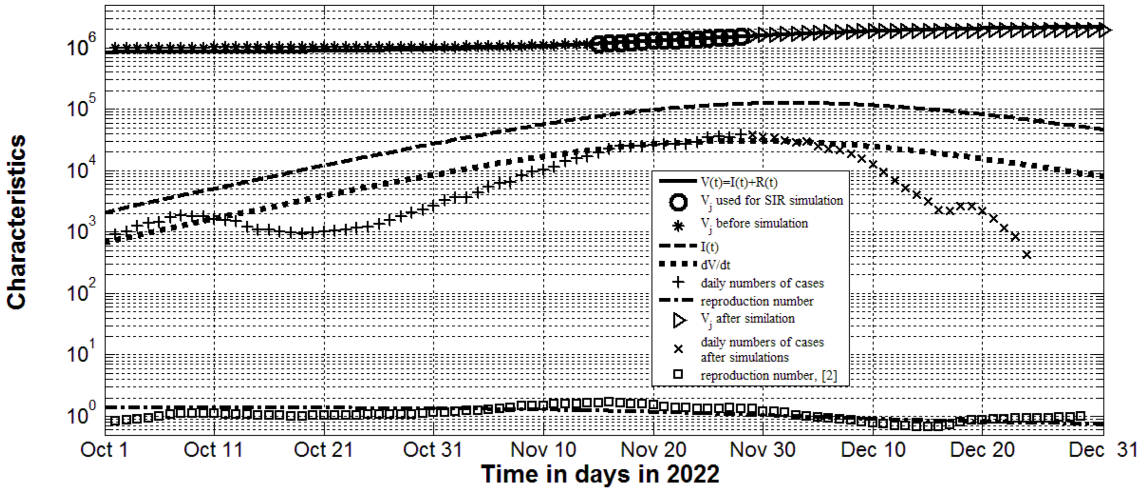

Table1. Cumulative numbers of laboratory-confirmed Covid-19 cases in mainland China for the period of October 1 to December 31, 2022 according to JHU report on January 4, 2023, [2].

Day in corresponding month of 2022	Accumulated numbers of cases, *Vj*		
	October	November	December
1	992046	1035560	1635593
2	992802	1038942	1698032
3	993657	1042353	1702449
4	994570	1046258	1733432
5	995650	1053804	1763679
6	996989	1054353	1788791
7	998779	1060280	1814055
8	1001871	1068278	1835148
9	1002305	1083133	1851983
10	1004398	1084571	1864738
11	1007872	1095429	1875371
12	1008241	1119690	1884218
13	1009732	1121325	1891352
14	1011191	1136846	1891352
15	1012384	1154441	1891352
16	1013390	1194415	1891352
17	1014291	1196687	1897331
18	1015212	1220590	1899290
19	1016148	1266052	1903956
20	1017089	1292056	1903961
21	1018065	1319652	1909905
22	1019038	1321605	1909905
23	1020028	1377221	1909905
24	1021117	1380570	1909905
25	1022331	1412498	1909905
26	1023743	1446896	1909905
27	1024984	1485399	1909905
28	1026239	1524446	1909905
29	1027864	1595756	1909905
30	1029918	1600201	1909905
31	1032790	-	1909905

The solid line show the number of victims *V(t)=I(t)+R(t)*, the dotted line – the theoretical estimations of the daily numbers of new cases *dV/dt*, the dashed line – the numbers of infectious persons *I(t)*, the dashed-dotted curve represent the effective reproduction numbers, calculated according to [8]. "Crosses" show the averaged daily numbers of new cases calculated with the use *Vj* values listed in Table 1 and formulas from [7, 8]."Stars", "circles", and "triangles" show the accumulated numbers of cases before, during and after SIR simulations, respectively. "Squares" show the values of the effective reproduction number from JHU dataset [2].

The registered accumulated numbers of cases (shown by "stars", "circles", and "triangles") are in very good agreement with the theoretical solid line. The smoothed registered daily numbers of cases (shown by "crosses") follow the theoretical dotted line. Therefore, the lifting of restrictions probably did not increase the number of new cases.The slightly lower numbers of daily cases (compare "crosses x" with the dotted theoretical curve) can be explained by a decrease in the level of testing and the fact that many cases were not registered.

The generalized SIR model allows estimating the effective reproduction number, which shows the average number of people infected by one person [8,10-14]. The corresponding dashed-dotted line (see Fig.1, [9]) are close to the vales calculated with the use of method proposed in [14], listed in [2] and shown by "squares". The reproduction numbers do not exceed 1.7 and are much lower than value 21 reported by the director of National Institute For Viral Disease Control and Prevention, China CDC, [15]. Probably this incredible high value of the reproduction number has been used by NHC to estimate the recent epidemic dynamics in China.

The calculated numbers of infectious people $I(t)$ make it possible to estimate the probability p of meeting a person spreading the infection, [7,8]. The maximum value 110,000 in early December, 2022 (see the dashed line) yielded the rather low value $p=7.6e\text{-}5$ for mainland China, [9]. This value is much lover, than the maximum probability 0.012 estimated for Japan August 2022,[8].

The given analysis allows us to conclude that the information about the millions of new Covid-19 cases appearing every day in China is improbable.

ACKNOWLEDGEMENTS

The author is grateful to Oleksii Rodionov for his help in collecting and processing data.

REFERENCES

1. Bloomberg news. Internet information. https://www.bloomberg.com/news/articles/2022-12-23/china-estimates-covid-surge-is-infecting-37-million-people-a-day. Accessed January 3, 2023.
2. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://github.com/owid/covid-19-data/tree/master/public/data. Accessed January 3, 2023.
3. Ryan M Barber, Reed J D Sorensen, David M Pigott, Catherine Bisignano, Austin Carter, Joanne O Amlag et al. Estimating global, regional, and national daily and cumulative infections with SARS-CoV-2 through Nov 14, 2021: a statistical analysis, The Lancet, Volume 399, Issue10344, 2022, pp.2351-2380, https://doi.org/10.1016/S0140-6736(22)00484-6.
4. I. Nesteruk, O. Rodionov. The COVID-19 pandemic in rich and poor countries. Preprint. Research Square. Posted December 7, 2022. https://doi.org/10.21203/rs.3.rs-2348206/v1.
5. Zero-COVID countries. Internet information https://en.wikipedia.org/wiki/Zero-COVID. Accessed January 3, 2023.
6. Internet information (in Chinese). https://news.cctv.com/2022/12/03/ARTIbCeNratl6uA4kXhzEdBG221203.shtml. Accessed January 3, 2023.
7. Nesteruk I. *COVID-19 pandemic dynamics*. Springer Nature. 2021. DOI: 10.1007/978-981-33-6416-5. https://link.springer.com/book/10.1007/978-981-33-6416-5.
8. Nesteruk I. Improvement of the software for modeling and forecasting the dynamics of epidemics and development of user-friendly interface. Preprint. Research gate, posted October 2022. DOI: 10.13140/RG.2.2.18942.66884.
9. Nesteruk I. The COVID-19 epidemic wave in mainland China at the end of 2022: monitoring and predicting with the use of the generalized SIR model. Preprint. Research Gate, posted December19, 2022. DOI:10.13140/RG.2.2.19479.24488.
10. Basic reproduction number. Internet information. https://en.wikipedia.org/wiki/Basic_reproduction_number.Accessed January 3, 2023.
11. Effective reproduction number. Internet information. https://www.r-bloggers.com/2020/04/effective-reproduction-number-estimation/. Accessed January 3, 2023.
12. an der Heiden, M., and O. Hamouda. 2020. "Schätzung Der Aktuel-Len Entwick lung Der Sars-Cov-2- Epidemie in DeutschLand – Nowcasting." *Epid-Bull* 17:10–15. https://doi.org/10.25646/669.
13. Cori, Anne, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. 2013. "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics." *American Journal of Epidemiology* 178 (9):1505–12. https://doi.org/10.1093/aje/kwt133.
14. Arroyo-Marioli F, Bullano F, Kucinskas S, Rondón-Moreno C (2021) Tracking R of COVID-19: A new real-time estimation using the Kalman filter. PLoSONE 16(1):e0244474. https://doi.org/10.1371/journal.pone.0244474.
15. Internet information of National Institute For Viral Disease Control and Prevention, China CDC (in Chinese). https://news.cctv.com/2022/11/28/ARTI7xtooQZazf58JAzgVjoE221128.shtml. Accessed January 3, 2023.