Case Fatality Ratio, Under-Reporting of Case Burden and Estimation of Excess Mortality: A Review of Challenges and Methods with Insights from COVID-19

Anuradha Rajkonwar Chetiya * , Priya Ranganathan , Vishal Deo

Review

Case Fatality Ratio, Under-Reporting of Case Burden and Estimation of Excess Mortality: A Review of Challenges and Methods with Insights from COVID-19

Anuradha Rajkonwar Chetiya [1],*, Priya Ranganathan [2] and Vishal Deo [3]

[1] Department of Statistics, Ramjas College, University of Delhi, Delhi, India

[2] Department of Anaesthesiology, Tata Memorial Centre, Homi Bhabha National Institute, Mumbai, Maharashtra, India

[3] National Institute for Research in Digital Health and Data Science, Indian Council of Medical Research, New Delhi, India

* Correspondence: anuradha_rc@hotmail.com

Abstract

Case fatality ratio (CFR) is a useful metric for determining severity of an epidemic in terms of mortality. This metric however needs to be used with caution. For the CFR to give an accurate estimate of the true impact of a pandemic, it must be based on the actual number of confirmed deaths and cases. However, significant under-reporting of the number of deaths and infected cases, especially during the early stages of the epidemic, makes it difficult to obtain a reliable estimate of CFR.Model based predictions based on such data can be subject to biases and could result in underestimation of the risk.It is thus imperative to ensure that the estimate of CFR accurately reflects the true nature of the epidemic.This paper discusses the problems linked with the estimation of CFR in the presence of under-reporting, and reviews some popular statistical methods which were used in the literature to estimate excess deaths due to COVID-19.

Keywords: case fatality rate; underreporting; epidemic; excess deaths

1. Introduction

Case fatality ratio (CFR) is a useful metric for determining severity of an epidemic in terms of mortality. CFR is defined as the ratio of the number of confirmed deaths to the number of confirmed cases of the disease during a given time period. Accurate estimation of CFR is important as it helps in analyzing the impact of a pandemic in multiple ways. It serves as a measure of mortality, or in other words, the potential of the disease outbreak to cause deaths. In addition, CFR also helps in assessing the true level of incidence and prevalence of a disease in any population.[1] Such insights during an epidemic are crucial for many public health related policy decisions, like identifying geographical areas and sub-populations which are at higher risk, planning and implementing preventive measures, resource allocation and assessing the preparedness of public health organizationsto efficiently deal with the disease outbreak. It is thus imperative to ensure that the estimate of CFR accurately reflects the true nature of the epidemic. However, significant under-reporting of the number of deaths and infected cases, especially during the early stages of the epidemic, makes it difficult to obtain a reliable estimate of CFR. [1,2]

CFR was used extensively across countries as a metric of measuring severity of the COVID-19 pandemic caused by the SARS-CoV-2 virus. However, significant under-reporting of the number of deaths and infected cases in several countries was reported by many studies. [3,4] So, to gauge the mortality risk associated with epidemics caused by novel viruses like SARS-CoV-2, we must resort

to some alternative analysis which can account for the under-reporting of cases. This paper discusses the problems linked with the estimation of CFR in the presence of under-reporting, and reviewssome popular statistical methods which were used in the literature to estimate excess deathsdue toCOVID-19.

2. Methods of Calculating CFR / Factors Affecting CFR Computation

Atkins et al.[2] have defined two methods to calculate the number of confirmed infections and number of confirmed deaths - one based on overall population-level data and the other based on individual patient-level data that traces the disease progression in every individual till the final outcome. Although population-level data is generally available even during the early phase of a pandemic, patient-level data, based on follow-up study, can be obtained at a relatively later stage of the pandemic. Of the two methods, the one based on patient-level data is considered to yield a more reliable estimate of the CFR, despite the late availability of data. During the COVID-19 pandemic, Verity et al. [5] conducted detailed analysis of individual patient-level data to obtain more robust estimates of the age-stratified CFRs. Atkins et al.[2] have defined four kinds of biases that may influence CFR calculation using population level data, one of them being under-estimation of deaths.

2.1. The Problem of Under Reporting and Estimation of CFR

Best estimates of CFR can be achieved by ensuring that the number of confirmed deaths and the number of confirmed cases are comprehensive and accurate. In other words, the same degree of reporting has been followed both for cases as well as deaths. In addition, since there is a lag between the time of reporting of a case and the final outcome of the case, and ideally, the CFR needs to be adjusted for such time lags. Wilson et al. [6]estimated CFRs of COVID-19 adjusted for time lag from reporting to death for four different populations including a group of 82 countries and a cruise ship.

According to the scientific brief of the World Health Organization (WHO), countries had different systems and methods of describing the numerator and the denominator of the CFR, that is, in reporting cases and deaths due to COVID-19. This resulted in biases in the estimation of CFR, and estimated ratios ranged from less than 0.1% to over 25% in different countries.[7] In other words, CFR estimates could be biased in either direction. Within the same country also, different states or jurisdictions may have had variable guidelines and methods of testing inconsistent definitions of data, ambiguous timing in reporting, disparity in data quality, and different mechanisms of ascertaining cases and deaths, see for example Galaitsi et al. [8], Vasudevan et al. [9].

Multiple factors may have led to the underestimation or underreporting of the true level of transmission in case of COVID-19. A substantial proportion of people with the infection went undetected, either because they were asymptomatic or had only mild symptoms and did not report at healthcare facilities.[10] Inadequate testing, either due to unavailability of testing kits during the initial period of the pandemic, or due to the strategy of testing only high-risk individuals also played a major role in underreporting. [1,3,4]In addition, method of testing also played a significant role in the level of detection. For example, the rapid antigen test for SARS-CoV-2 is known to have lower sensitivity than RT-PCR, and its sensitivity further reduces in asymptomatic individuals.[11] Presence of a high proportion of individuals with mild or no symptoms among COVID-19 infected individuals demanded increased scale of random tests to get the accurate number of confirmed cases. Despite repeated appeals and advisories to all countries from the WHO to employ extensive random testing, only few countries could conduct an adequate number of COVID-19 tests.[1] Testing of focused groups could only provide information about the incidence of infection among the high-risk fraction of the susceptible. Under-reported value of the denominator may result in the CFR value being overestimated, as was seen in the case of the UK which had one of the highest case fatalities in the world. The actual number of infected cases in the UK was estimated to be about eight times more than the number of reported cases (http://covid.econ.cam.ac.uk). Intensity and timing of increases in testing during the course of the Covid-19 pandemic have been found to be associated with such

discrepancies.[12]More research studies on the problem of under-reporting can be found in the works of Gamado et al. [13], Prado et al. [14] and Krantz and Rao[15].

If the percentage of under-reporting is same in the case of both confirmed cases and reported deaths, the CFR based on individual patient level data is expected to give an estimate more or less similar to the one based on reported population level data. However, if the level of under-reporting is not the same in confirmed cases and deaths, this assumption will not hold true. In such situations, a possible way out is estimating the excess deaths associated with COVID-19. Estimating excess deaths is very useful in evaluating the true impact of any ongoing pandemic.   The Centers for Disease Control and Prevention (CDC)defined excess deaths due to COVID-19 as the difference between the observed numbers of deaths in specific time periods and expected numbers of deaths in the same time periods.[16]

The actual number of deaths due to COVID -19 could be under-reported due to different reasons. The common cause in most countries was, again, a lack of comprehensive testing. Another reason being the differences in the manner in which countries define deaths due to COVID-19. As a result of which, deaths attributed to some other diseases with similar clinical symptoms may actually be due to COVID-19. Monitoring mortality from all causes and diseases like influenza or pneumonia with symptoms similar to COVID-19 can help bridge the gap between reported and actual deaths attributed to this virus. The CDC has estimated excess death counts by observing the overall all-cause mortality, that is deaths due to all causes, and all-cause mortality excluding COVID -19 and then using their estimates to understand how many excess deaths can be attributed to COVID-19.[16] A brief review of methods used for estimating excess mortality due to COVID-19 is presented in the subsequent section.

3. Estimation of Excess Mortality Due to COVID-19: A Review of Methods

One of the pioneers of the use of excess mortality to measure the level of transmission of any epidemic was Dr. Selwyn Collins. He used excess deaths to conduct detailed studies of epidemiological characteristics of the influenza outbreak, one of them being a groundbreaking 35 year analysis of data to estimate excess annual deaths. Serfling [17] defined three criteria for estimation of excess deaths from pneumonia- influenza epidemic. They are, determination of secular trend and its extrapolation, estimation of seasonal variation, and distinction between epidemiologically significant departures from expected weekly mortality and random variation from endemic levels.   The models for estimation of excess deaths due to COVID -19 epidemic reviewed in this section fulfill the essence of these requirements, though the methods used to do so differ somewhat from the ones used by Serfling[17].

Weinberger et al. [12] have demonstrated that estimates of the death toll of COVID19 based on excess all-cause mortality may be more reliable than those relying only on just reported deaths, particularly in places that lack widespread testing.Their study, based on different states across the United States (US),firstestimated the baseline number of deaths in the absence of COVID-19. This estimate was then subtracted from the observed number of deaths for the period from March 1, 2020, to May 30, 2020. To obtain the baseline estimate, the authors used Poisson regression models fitted to the weekly state level death counts for five years from 2015 to 2020. This was then projected forward to generate current baseline deaths. The study also adjusted the baseline model for seasonality, year-to-year baseline variation, influenza epidemics, and reporting delays. They conclude that estimation of excess deaths is important as this provides an estimate of the full COVID-19 burden. The study also found that the official tallies of deaths due to the virus are under-reported. Overall 122300 excess all-cause deaths were estimated for the US during the study period out of which only 78% were officially attributed to COVID -19 and the remaining 22 % were unattributed to COVID-19. They also found that the proportion of excess deaths that were attributed to COVID-19 varied between states and increased over time.

Wetzler and Cobb [18] conducted a follow up study to determine how many out of the remaining 22% deaths from the results of the research by Weinberger et al. [12] could be attributed to COVID -

19. CDC reports weekly data of the 13 causes of death from the most prevalent comorbid conditions reported on death certificates where COVID-19 was listed as a cause of death. The 2015-2019 data for weeks 10 through 22 were used in this study to forecast the number of deaths from the 13 causes in the absence of COVID-19 during 2020. The forecast was subtracted from the observed number of deaths for each cause during the period March 1 to May 30, 2020. Based on their findings they concluded that as much as 93% of the excess deaths were due to COVID -19.

This result is similar to the study by Wetzler and Wetzler [19], who estimated the percentage of excess deaths in the US in the month of April to be 95%, although the methods used in both studies are different. Wetzler and Wetzler[19] used the Attributable Fraction Method introduced by Levin in 1953, to estimate the number and percentage of excess deaths. This was then further used to estimate the number of months it would take for these excess deaths to occur under normal circumstances, the incidence of infection, and under-detection of infection. Attributable fractions (AFs) are measures of association between a disease and a specific exposure that attempt to assess the public health impact of that exposure. (as defined in www.publichealth.columbia.edu/research/population-health-methods/adjusted-attributable-fractions).

Riveria et al.[20] estimated excess all-cause, pneumonia, and influenza mortality during COVID-19 using two models viz. a semi-parametric model and a conventional time-series model. The analysis was based on data from 9 states of the US with high reported COVID-19 deaths and reportedly complete mortality data. Their study used a general additive model with the covariates and a quasi-Poisson model to estimate the dispersion parameter. The conventional excess mortality method consists of first fitting a quasi-Poisson semi-parametric model with expected rate being the mean weekly deaths during this period. Weekly deaths are approximated by a normal distribution with a variance that accounts for over-dispersion.   The study concluded that results obtained from both the models were consistent although results from the semi-parametric model were found to be less precise.

A cohort study in the UK by Banerjee et al.[21] provides initial estimates of the excess COVID-19-related deaths over a one-year period based on four different rates of infection, by using Kaplan Meier time-to-event analysis. Chakravarty[22] in a study using infection fatality rates (IFR) of different age groups estimated missing deaths in COVID-19 records in Delhi, to be in the range of 1500- 2000 in the age group 60 years or more.

Sweden was perhaps the only country in the world with a different approach to the pandemic than other countries. No restrictions were imposed by the Swedish government to contain the spread of the virus, the objective being to attain herd immunity. However, it turns out that Sweden is among the few countries in Europe with highest reported excess mortality (https://euromomo.eu/). In a study by Modig and Ebeling [23], weekly age- and sex-specific death rates of data from Sweden for five years were analyzed to obtain more accurate estimates of the excess mortality attributed to COVID-19.The results from this study indicate higher death rates than previous years in all age groups over 60 years in Sweden. In the age group 80 years and above, men suffered higher levels of excess mortality than women at all ages with 75% higher death rates for males and 50% higher for females.

A study by the COVID-19 excess mortality collaborators, Wang et. al. [24], sourced data from 74 countries and territories, as well as 252 subnational units that reported weekly or monthly all-cause mortality during 2020-21. Excess mortality was calculated as observed deaths minus expected deaths, after adjusting for late reporting and excluding periods with anomalies (e.g., heatwaves). The expected mortality was estimated using an ensemble of six models, weighted based on their predictive accuracy. For locations lacking direct mortality data, the study used least absolute shrinkage and selection operator (LASSO) regression with 15 covariates, including COVID-19-related factors (e.g., seroprevalence) and general health metrics. The study estimated that 18.2 million people died globally due to the COVID-19 pandemic between January 1, 2020, and December 31, 2021. This figure is3.07 times higher than the 5.94 million reported COVID-19-related deaths. The highest excess mortality rates were observed in Andean Latin America, Eastern Europe, and CentralEurope**.**

Conversely, the lowest excess mortality rates were reported in East Asia, Australia**,** and high-income Asia Pacificcountries.

The study highlighted significant discrepancies between reported COVID-19 deaths and actual excess mortality, particularly in regions with limited testing and under-reporting, such as sub-Saharan Africa and Central Asia. The findings indicate that the pandemic's full mortality impact was significantly underreported in many countries, due to insufficient testing and reporting challenges. The study emphasized the importance of using excess mortality estimates to better understand the pandemic's true impact, especially in regions where under-reporting and data gaps were common.

The WHO formed a Technical Advisory Group (TAG) of experts across demography, epidemiology, and statistics to develop methods for estimating excess mortality globally.[25] Data was taken from country focal points, public databases, and mortality datasets. Over 140 countries participated in this consultation process, which involved reviewing draft estimates and providing additional data. However, many countries lacked complete all-cause mortality (ACM) data. For these countries, WHO used a Poisson count model to predict monthly deaths, filling gaps where necessary. Bayesian Poisson framework was used to estimate both the expected deaths (for all countries and all months) in the absence of the pandemic, and the ACM for countries where it was not available during the pandemic. Variables with spatio-temporal variations such as COVID-19 test positivity rate, containment measures, average national temperature, historic cardiovascular disease death rates, income-level of country, etc., considered to be associated with changes in excess mortality over the course of the pandemic, were used in the models. In the locations without adequate reporting of mortality, a log-linear regression model within the Bayesian Poisson framework was used to predict mortality levels. For a few countries, their sub-national observed deaths were used to predict the national deaths using multinomial models. This was done under the assumption that the relationships estimated between sub-national and national mortality estimates persisted during the pandemic as well. ACM data were aggregated by year or sub-national regions, and models predicted expected mortality using historical and pandemic-period data. The reported or predicted deaths and the expected deaths (in a non-pandemic situation) were used to estimate monthly excess deaths in all locations for 2020 and 2021.

WHO estimated that 14.83 million excess deaths occurred globally due to the COVID-19 pandemic in 2020 - 2021. The study revealed that excess mortality in 2021 was significantly higher than in 2020, driven by continued waves of infections and deaths. The global excess mortality was found to be 2.74 times higher than the number of reported COVID-19 deaths. Many countries, particularly those with weaker health systems and limited reporting infrastructure, underreported the true impact of the pandemic. The WHO emphasized the importance of accurate excess mortality data in forming future health policies. Following the guidelines for Accurate and Transparent Health Estimates Reporting, the estimates are periodically refined as more data becomes available.

4. Discussion

Although the extent of under-reporting may vary between settings, under-reporting of cases is a global problem. This is evident from the current review of research on excess deaths related to COVID-19, across different countries, which have shown this to be a substantial problem. Given such situations, estimation of excess deaths provides a scientifically rigorous method of assessing true mortality risk associated with an epidemic. Apart from the Poisson regression models, various other statistical and mathematical models like Kaplan Meier time-to-event analysis, Attributable Fraction Method, Age specific infection fatality rates have been used by researchers to estimate excess mortality. These methods are essential to be used especially during the early stages of the epidemic so that comprehensive health and economic policies can be framed at the earliest to face the grave challenges presented by the epidemic.

Competing interests statement: The authors declare that they have no competing interests

Patient and Public Involvement (PPI) Statement: Patients and the public were NOT INVOLVED in the design, or conduct or reporting or dissemination plans of our research.

References

1. Deo V, Grover G. A new extension of state-space SIR model to account for underreporting–an application to the COVID-19 transmission in California and Florida. Results in Physics. 2021 May 1;24:104182.
2. Atkins KE, Wenzel NS, Ndeffo-Mbah M, Altice FL, Townsend JP, Galvani AP. Under-reporting and case fatality estimates for emerging epidemics. bmj. 2015 Mar 16;350.
3. Wu SL, Mertens AN, Crider YS, Nguyen A, Pokpongkiat NN, Djajadi S, Seth A, Hsiang MS, Colford Jr JM, Reingold A, Arnold BF. Substantial underestimation of SARS-CoV-2 infection in the United States. Nature communications. 2020 Sep 9;11(1):4507.
4. Lau H, Khosrawipour T, Kocbach P, Ichii H, Bania J, Khosrawipour V. Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters. Pulmonology. 2021 Mar 1;27(2):110-5.
5. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, Cuomo-Dannenburg G, Thompson H, Walker PG, Fu H, Dighe A. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet infectious diseases. 2020 Jun 1;20(6):669-77.
6. Wilson N, Kvalsvig A, Barnard LT, Baker MG. Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality. Emerging infectious diseases. 2020 Jun;26(6):1339
7. World Health Organization. Estimating mortality from COVID-19: Scientific brief, 4 August 2020. World Health Organization; 2020.
8. Galaitsi SE, Cegan JC, Volk K, Joyner M, Trump BD, Linkov I. The challenges of data usage for the United States' COVID-19 response. International Journal of Information Management. 2021 Aug 1;59:102352.
9. Vasudevan V, Gnanasekaran A, Sankar V, Vasudevan SA, Zou J. Disparity in the quality of COVID-19 data reporting across India. BMC public health. 2021 Jun 24;21(1):1211.
10. Byambasuren O, Cardona M, Bell K, Clark J, McLaws ML, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. Official Journal of the Association of Medical Microbiology and Infectious Disease Canada. 2020 Dec 11;5(4):223-34.
11. Peña M, Ampuero M, Garcés C, Gaggero A, García P, Velasquez MS, Luza R, Alvarez P, Paredes F, Acevedo J, Farfán MJ. Performance of SARS-CoV-2 rapid antigen test compared with real-time RT-PCR in asymptomatic individuals. International Journal of Infectious Diseases. 2021 Jun 1;107:201-4.
12. Weinberger DM, Chen J, Cohen T, Crawford FW, Mostashari F, Olson D, Pitzer VE, Reich NG, Russi M, Simonsen L, Watkins A. Estimation of excess deaths associated with the COVID-19 pandemic in the United States, March to May 2020. JAMA internal medicine. 2020 Oct 1;180(10):1336-44.
13. Gamado KM, Streftaris G, Zachary S. Modelling under-reporting in epidemics. Journal of mathematical biology. 2014 Sep;69:737-65.
14. Prado MF, Antunes BB, Bastos LD, Peres IT, Silva AD, Dantas LF, Baião FA, Maçaira P, Hamacher S, Bozza FA. Analysis of COVID-19 underreporting in Brazil. Brazilian Journal of Intensive Care. 2020 Jun 24;32:224-8.
15. Krantz SG, Rao AS. Level of underreporting including underdiagnosis before the first peak of COVID-19 in various countries: Preliminary retrospective results based on wavelets and deterministic modeling. Infection Control & Hospital Epidemiology. 2020 Jul;41(7):857-9.
16. CDC. (2020). Excess Deaths Associated with COVID-19. Retrieved September 10, 2024, fromhttps://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm.
17. Serfling RE. Methods for current statistical analysis of excess pneumonia-influenza deaths. Public health reports. 1963 Jun;78(6):494.
18. Wetzler HP, Cobb HW. New insights on excess deaths and COVID-19. medRxiv. 2020 Jul 27:2020-07.

19. Wetzler H, Wetzler E. COVID-19 Excess Deaths in the United States, New York City, and Michigan During April 2020. medRxiv. 2020 Apr 6:2020-04.

20. Rivera R, Rosenbaum JE, Quispe W. Excess mortality in the United States during the first three months of the COVID-19 pandemic. Epidemiology & Infection. 2020 Jan;148:e264.

21. Banerjee A, Pasea L, Harris S, Gonzalez-Izquierdo A, Torralbo A, Shallcross L, Noursadeghi M, Pillay D, Sebire N, Holmes C, Pagel C. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. The Lancet. 2020 May 30;395(10238):1715-25.

22. Chakravarty S. Estimating missing deaths in Delhi's COVID-19 data. MedRxiv. 2020 Jul 30:2020-07.

23. Modig K, Ahlbom A, Ebeling M. Excess mortality from COVID-19: weekly excess death rates by age and sex for Sweden and its most affected region. European journal of public health. 2021 Feb 1;31(1):17-22.

24. Wang H, Paulson KR, Pease SA, Watson S, Comfort H, Zheng P, Aravkin AY, Bisignano C, Barber RM, Alam T, Fuller JE. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. The Lancet. 2022 Apr 16;399(10334):1513-36.

25. Msemburi W, Karlinsky A, Knutson V, Aleshin-Guendel S, Chatterji S, Wakefield J. The WHO estimates of excess mortality associated with the COVID-19 pandemic. Nature. 2023 Jan 5;613(7942):130-7.