Institute
and Faculty
of Actuaries

COVID-19: Statistics, Ratios and understanding limitations of data available

ICAT: Data Limitations sub-group

L.R. Lopes Junior, S. Aggarwal, M. Qazvini

This report is produced by a Health and Care workstream of ICAT (IFoA Covid-19 Action Taskforce)

12 December 2020

Introduction

This section reviews common statistics and ratios about the spread of the virus among humans during the ongoing COVID-19 pandemic. As a result of the evolving nature of the virus and changing dynamics of ways to tackle it by respective governments, there are several limitations in the use of such statistics and ratios. Lack of reliable data and the vast amount of information make it difficult to determine the reliability of different resources.

This report is expected to increase awareness of readers not only about the definitions of measures used but also about the limitations around the available data. Further, the report highlights an increase in the need for readers to be cautious while using different information and statistics in forming a view of the extent of the reach of the virus among masses.

Confirmed Cases

Number of confirmed cases is, generally, the count of people who have tested positive by polymerase chain reaction (PCR) testing. PCR tests for the presence of the genetic material of the virus.

The number of confirmed cases is generally much lower than actual number of people that have been infected with COVID-19 for several reasons:

- Some proportion of those infected are not symptomatic or mildly symptomatic such that they never seek medical care and hence may not be tested. (1)
- PCR testing only picks up the virus in the blood stream of the infected individual. Thus, if infected individuals are tested long after the infection day, the results of the PCR test are likely to be negative.
- Over the course of the epidemic the testing regimes in a particular country or region may have been altered and this results in changes in confirmed COVID-19 cases over time. (2)
- Similarly, different countries or regions within a country may have different capabilities about testing and hence this may affect the numbers of confirmed cases. While some countries are undertaking a massive scale of testing (e.g. drive through testing in South Korea), many developing countries are still struggling to ramp up their testing capacity.

Based on the above it's clear that comparisons of case numbers over time and between different regions need to consider differences in practices with regards to testing.

A further consideration should be that the data may be delayed from when the test was made to when the test result came back, especially if there are testing backlogs.

Infections

This relates to the true number of people who are infected, which includes confirmed cases, asymptomatic cases (that show no or limited symptoms) and symptomatic cases not tested. Since the true rate of infection could only be known if everyone was tested, this statistic is not seen often quoted and instead must be estimated. For estimation, models which calibrate assumptions around the number of confirmed cases or the number of deaths or a mix of both can be used. Seroprevalence (the level of a pathogen in a population, as measured in blood serum) studies such as those conducted in the UK (3) also try to estimate the number of infections in a population.

One limitation of models trying to estimate the infection rate is that the proportions of asymptomatic cases and symptomatic cases not tested in a population are not observed. Thus, these proportions need to be estimated, carrying some error level, which will likely impact infection rate modelling.

Reported COVID-19 Deaths

Generally, governments report (often on a daily basis) the number of deaths from COVID-19. These tend to be deaths from confirmed cases or sometimes even deaths that were posthumously tested. Some countries may also include suspected COVID-19 deaths in their counts, for example if people died that had symptoms associated with COVID-19.

As described above for cases, these figures are sensitive to practices in different countries and even regions within the country, for example testing issues. It's also potentially subject to reporting delays.

In the UK an example of reporting variations was that initially the epidemic deaths in hospitals were the primary source of reported deaths, and the figures missed a lot of deaths in care homes, and also in the home (4).

A further example can be seen in Florida where there recently appeared to be a downturn in the number of deaths occurring, but in fact there was no downturn, just a delay in the number of reported deaths (5).

Excess deaths

The cumulative deaths related to COVID-19 provides only a partial view of the impact of the pandemic in overall mortality. According to the most recent mortality monitor report from the Continuous Mortality Investigation (6), there have been 62,100 excess deaths in UK from the start of the pandemic to week 30 of 2020 (i.e. week starting on 24 July 2020).

The excess deaths are estimated as the difference between the expected number of deaths and the observed number of deaths. The expected number of deaths is based on the Standardised Mortality Rate (SMR) experienced in the same period of 2019. In other words, it represents the number of deaths that would be observed if the SMR in 2020 was the same as observed in 2019.

Care must be taken when attributing this excess of deaths to COVID-19 only. As mentioned in the CMI report (6), some deaths related to COVID-19 may not be considered excess because the death could have occurred in the same period, but from another cause. Also, the pandemic might have accelerated the occurrence of deaths, a phenomenon called "forward mortality displacement". The limitations mentioned previously in this report, regarding the registration of deaths and attribution of the cause to COVID-19 also apply here.

To help illustrate the evolution of the SMR in 2020, the graph in Figure 1 shows that the cumulative mortality in 2020 was similar to that in 2019 up to week 12 (before the pandemic hit UK). In week 30, it can be observed that it is 6.4% higher than the 2010-2019 average. However, its peak was in week 23, and since then, the SMR has been decreasing. In fact, weeks 25 to 30 registered lower mortality than in the same period of 2019. Only by the end of 2020, the full picture of the impact in mortality caused by COVID-19 in overall mortality can be observed.

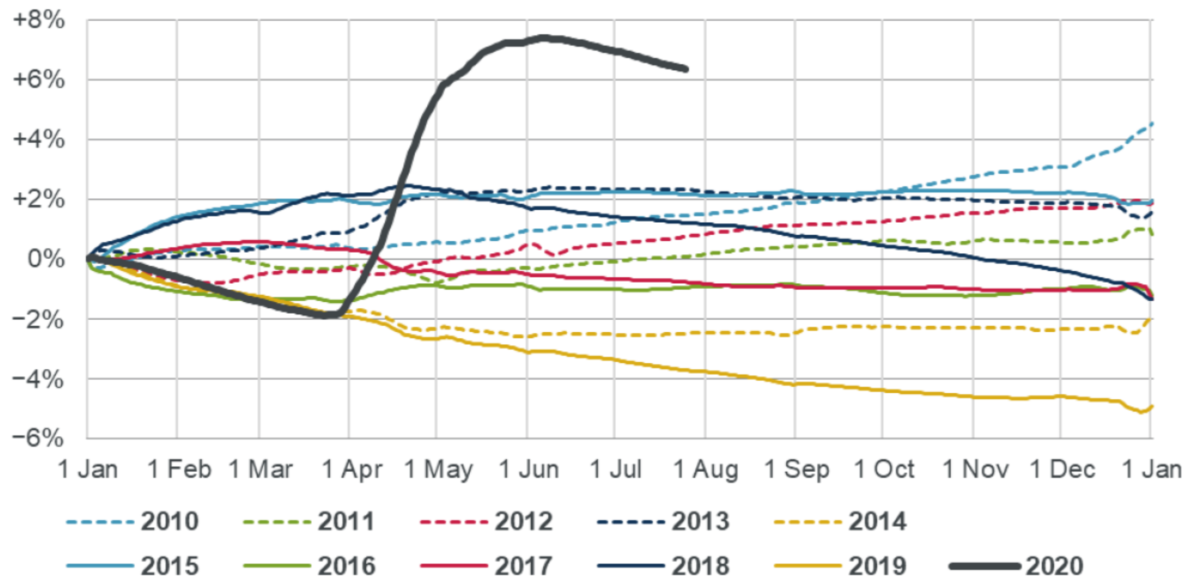

Figure 1: Cumulative standardised mortality rate compared to the 2010-2019 average. Source: Mortality monitor - COVID-19 update - week 30 of 2020, CMI 2020.

Insurers are concerned about the impact of the epidemic on its overall deaths count and not just COVID-19 deaths.

While pandemic-induced health system failure could lead to rise in mortality from non-COVID-19 deaths, the lockdown would cause a reduction in deaths from motor or industrial accidents. It, therefore, becomes imperative to understand emerging data from different sources. For example, the UK government has only focussed on hospital COVID-19 deaths in their reporting thereby excluding COVID-19 deaths outside hospitals and non-COVID-19 deaths.

In order to consider overall mortality, regardless of the case, the reports published by the Continuous Mortality Investigation that compare overall deaths during the pandemic with the deaths in the same period of previous years is a good source. (6)

Case Fatality Rate (CFR)

$$Case\ Fatality\ Rate\ = \frac{Number\ of\ confirmed\ deaths}{Number\ of\ confirmed\ cases}$$

CFR measures the proportion of deaths over the confirmed number of cases during a specified period, and is an indication of the severity of a disease (7). It is mainly used for acute conditions, with short duration, rather than chronic diseases with long durations. This is a highly published measure and can be easily found in dashboards that compare the pandemic data by country.

During an outbreak, the CFR can be a poor measure of the mortality of a disease for the following reasons:

- At a given point in time during an epidemic, not all cases have recovered or died, some lives are still “sick” and so the CFR can be too high or too low as the epidemic progresses. For example, where case numbers are increasing rapidly the CFR would be lower than the true CFR. The true CFR can only be calculated once all cases are resolved (recovered or died).

- Comparisons between countries and over time should also be made with care, as testing approaches are different depending on the country and may also change over time.
- Similarly, the reporting with regards to deaths could have changed.
- The CFR of COVID-19 also varies by other factors such as age, gender, smoking habits and presence of co-morbidities, as they impact the incidence of cases and deaths (2). Thus, any comparison of a single CFR may carry less meaning when comparing across populations with different mixes of these factors. CFRs can be calculated by age and other risk factors.

Infection Fatality Ratio (IFR)

$$Infection\ Fatality\ Ratio = \frac{\begin{array}{c}Total\ number\ of\ deaths\\ from\ a\ specific\ disease\end{array}}{Total\ number\ of\ infections}$$

IFR measures the proportion of all deaths from a specific disease over the total number of infected individuals. Thus, IFR is much lower than CFR. For instance, the estimated CFR of COVID-19 in China is 1.38% while the estimated IFR is 0.66% (8). This is an important indicator, as policy makers are more interested in making decisions based on burden of the disease in the whole infected population rather than confirmed cases only (9).

As the number of infections and, in some cases, also accurate numbers of deaths are unknown, the infection fatality ratio is difficult to estimate for COVID-19 The high asymptomatic rate of the disease is one of the main contributory reasons for the difficulty. To improve the estimate the infection fatality rate various approaches are possible that may involve some modelling and/or seroprevalence studies that test for antibodies for SARS-CoV-2 (to estimate the number of infections).

Based on the above some limitations are apparent:

- There is usually uncertainty in the estimates of the IFR because there will be uncertainty in both the estimates for infections and the modelling approach used to derive the link between infection and deaths.
- IFR also varies strongly by age and other factors making a single number less meaningful when comparing populations with different age profiles. IFRs are calculated by age bands for this reason.
- Only over time will the true IFR become easier to calculate.

Attack Rate

$$Attack\ Rate = \frac{Number\ of\ cases}{Population\ at\ risk}$$

The attack rate is calculated as the number of people who became ill divided by the number of people at risk for the illness (10). It is sometimes referred to as incidence proportion (11) and typically used for acute conditions rather than chronic.

Definitions need to be established for what would be considered a case and what would be the population at risk. For COVID-19, the population at risk can be the entire population, since there may be no known natural immunity to the disease.

One can consider cases as before (i.e. those testing positive) and define a case attack rate:

$$Case\ Attack\ Rate(CAR) = \frac{Confirmed\ cases}{Population\ at\ risk}$$

CAR considers only new cases confirmed by tests, which carries the same biases as the denominator of CFR.

Alternatively, by substituting confirmed cases with total number of new infections during a specific time, an Infection Attack Rate (IAR) can be defined as:

$$Infection\ Attack\ Rate(IAR) = \frac{New\ Infections}{Population\ at\ risk}$$

IAR considers all new infections (confirmed cases, asymptomatic, and non-tested). Thus, the main limitation of this measure is that its numerator is based purely on estimations.

The attack rate of SARS-CoV-2 can only currently be estimated from modelling and even at the end of the epidemic the attack rate will remain an estimate and the exact number of infections will (most likely) never be known.

The Reproduction Number

The basic reproduction number, often denoted with R_0, represents the average number of infections generated from one infected individual, in the absence of intervention, immunity from previous infections or vaccination. In other words, it is the average number of people infected by a single infected person before any response to the disease (all individuals are exposed to the disease). Numbers larger than 1 mean that the growth of the infections is exponential.

One would be tempted to think that there is a single basic reproductive number for SARS-CoV-2 but, it probably depends upon many factors. For example, it is likely to be different in a densely populated city versus the countryside. Similarly, many other factors such as household size and composition, local habits and other factors such as temperature and humidity may impact the spread of a virus.

Estimates for R_0 for SARS-CoV-2 average estimates of peer-reviewed studies range from around 1 to around 4, with an average of 2.54 (12)

The effective reproduction number (R_t) is the infectivity over time as society responds to epidemic. In the case of no interventions or change in behaviour, it can be found from the product between R_0 and the proportion of the population that is still susceptible to the disease at a given point in time. Thus, all else being equal the effective reproductive number is expected to reduce over time.

In practice though interventions such as lockdowns, mask wearing and changes in behaviour contribute to changes to the effective reproductive number over time.

Herd immunity threshold is an estimate of the proportion of the population that need to be infected with the virus which would result in a slowdown of the spread of the virus. This is usually calculated as $1 - {}^{1}/_{R_0}$. Once this proportion of the population has been infected each infection will generate less than one new infection on average and the virus should slowly stop spreading until it stops. The concept is called herd immunity because as a certain proportion of the population is infected the whole population essentially ceases to spread the disease effectively. Note however that the virus would still infect more than the herd immunity threshold as it will continue to spread until it stops. This additional spread is called the overshoot.

Both reproductive numbers are estimated using statistical/mathematical models that involve uncertainty from assumptions and parameter estimates.

Recent epidemiological and medical research

Since the start of the COVID-19 pandemic the number of resources and available information have grown rapidly and thanks to advanced technology, researchers from all over the world have been able to share their findings through virtual conferences and journals making the review process of COVID-19 related topics eight times faster than other topics. Such a quick response in scientific communities was unprecedented and not comparable with previous pandemics.

According to the University of Pittsburgh Medical Centre, between 1 January 2020 and 16 May 2020, in just 16 journals, 294 COVID-19 related articles were published. However, with this flow of information, it is not so easy to identify reliable resources and there are many works being published without being peer-reviewed. Recently, MIT Press launched a new journal, "Rapid Review: COVID-19" to review preprints articles in SSRN, bioRxiv, etc (13). In January 2020, the Welcome Trust in London called for researchers and journals to share their outputs "rapidly and openly". There are many journals that provide COVID-19 related papers without subscription fees (14).

On the other side, the consumers of COVID-19 information are also increasing fast with COVID-19 papers being downloaded more than 150 million times. But can all this information be trusted? Here are some tips (15):

- Make sure the sources of information are reliable. For example, that they are provided by well-known journals, magazines and research centres.
- Make sure providers of the information have required expertise; For examples epidemiologist, modellers with a strong research background.
- Compare different resources.
- Make sure the information provided is transparent. The more transparent models and data are, the more confident are the providers in their results.

Conclusions

As the pandemic unfolds, we are faced with the task to understand and apply data and concepts previously unknown by many of us. This report focused on defining and clarifying the information most published in the news and studies about COVID-19. It is not only important to comprehend what the number of cases, infections and deaths truly represent, but also to be aware of existing limitations involved in gathering and analysing these numbers.

Additionally, we covered many indexes used by experts to measure the severity of the disease and how fast the virus spreads in a population. Although they are very useful to determine public policies (such as the duration of a lockdown) and to set strategies to fight the pandemic, they need to be analysed and used with care in order to avoid misleading conclusions. We highlighted the main limitations and dangers when using these measures, especially if the aim is to compare the morbidity and mortality of the disease in different countries.

References

1. *Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020.* **Mizumoto, Kenji, et al.** s.l.

2. **Oke, Jason and Heneghan, Carl.** Global Covid-19 Case Fatality Rates. *Centre for Evidence-Based Medicine.* [Online] 09 June 2020

3. **Public Health England.** Weekly Coronavirus Disease 2019 (COVID-19) Surveillance Report. *National COVID-19 surveillance reports.* [Online] 07 August 2020. [Cited: 10 August 2020].

4. **Office for National Statistics.** Deaths involving COVID-19 in the care sector, England and Wales: deaths occurring up to 12 June 2020 and registered up to 20 June 2020 (provisional). [Online] 03 July 2020. [Cited: 08 August 2020].

5. **Glassman, Rebecca and Lacan, Olivier.** Florida's COVID-19 Data: What We Know, What's Wrong, and What's Missing. [Online] The COVID Tracking Project, 10 July 2020. [Cited: 08 August 2020].

6. **CMI.** *Mortality monitor – COVID-19 update – week 30 of 2020.* s.l. : Institute and Faculty of Actuaries, 2020.

7. **Harrington, Rebecca A.** Case fatality rate. *Encyclopædia Britannica.* [Online] 05 May 2020.

8. *Estimates of the severity of COVID-19 disease.* **Verity, Robert , et al.** 2020, MedRxiv.

9. *A systematic review and meta-analysis of published research data on COVID-19 infection-fatality rates.* **Meyerowitz-Katz, Gideon and Merone, Lea.** 2020, MedRxiv.

10. **Pettygrove, Sydney.** Attack rate, epidemiology. *Encyclopædia Britannica.* [Online] 15 March 2016.

11. **Dicker, Richard C., et al.** Morbidity Frequency Measures. *Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics.* 2006, pp. 3-11.

12. *Early in the epidemic: impact of preprints on global discourse about COVID-19 transmissibility.* **Majumder, Maimuna S and Mandl, Kenneth D.** 2020, The Lancet Global Health, pp. e627--e630.

13. **McKenzie, Lindsay.** Debunking bad COVID-19 research. [Online] INSIDE HIGHER ED, 20 June 2020.

14. **Nature Index.** COVID-19 research update: Original research findings crowded out by commentary. [Online] Nature Index, 2020.

15. **Balsari, Satchit, Buckee, Caroline and Khanna, Tarun.** Which Covid-19 Data Can You Trust? [Online] Harvard Business Review, 08 May 2020.

Institute and Faculty of Actuaries

London

7th Floor · Holborn Gate · 326-330 High Holborn · London · WC1V 7PP
Tel: +44 (0) 20 7632 2100 · Fax: +44 (0) 20 7632 2111

Edinburgh

Level 2 ·Exchange Crescent · 7 Conference Square · Edinburgh ·EH3 8RA
Tel: +44 (0) 131 240 1300 · Fax +44 (0) 131 240 1311

Oxford

1st Floor · Park Central · 40/41 Park End Street · Oxford · OX1 1JD
Tel: +44 (0) 1865 268 200 · Fax: +44 (0) 1865 268 211

Beijing

6/F · Tower 2 · Prosper Centre ·5 Guanghua Road · Chaoyang District · Beijing · China 1000020
Tel: +86 (10) 8573 1000

Hong Kong

2202 Tower Two · Lippo Centre · 89 Queensway · Hong Kong
Tel: +11 (0) 852 2147 9418 · Fax: +11 (0) 852 2147 2497

Singapore

163 Tras Street · #07-05 Lian Huat Building · Singapore · 079024
Tel: +65 6717 2955

www.actuaries.org.uk