The COVID-19 pandemic in Australia brought into focus an area that has typically seen little public interest or attention by individuals: the collection, analysis, and portrayal of health data – in this case, metrics related to COVID-19. There was overwhelming demand for this information - both initially and ongoing - to allow the public to educate themselves, and to track outbreaks and the overall progress of the pandemic. Much of the presentation and availability of data throughout the pandemic has relied on volunteers and unofficial sources.

From late 2021 and continuing to this day, I began collecting and reporting on COVID-19 metrics across jurisdictions in Australia, after concluding my earlier work in mapping COVID-19 outbreaks (all published under the handle [redacted] on the Twitter platform). By that time, there were already several non-official sites dedicated to the collection of these metrics, including Covid Live (https://covidlive.com.au/), Covid Base AU (https://covidbaseau.com/) and Covid 19 Data (https://www.covid19data.com.au/). However, differences in what metrics these sites collected, and differences in how they calculated others, as well as a focus on raw numbers at the time (with the exception of Covid 19 Data), led me to start posting visualisations and overviews of the Covid data in Australia, as I felt that these – along with numeric data for reference – gave a better glimpse of how the pandemic was evolving compared to the then-ubiquitous daily dashboards from the states and territories. My work has involved collaboration with the creators of those sites and others in the COVID-19 reporting community and has been utilised or referenced by them and in many different media articles also [1].

It was through this work that I started to realise the scope of differences across jurisdictions, and just how much reporting relied on assumptions of equivalence and missed crucial, though complicated, nuances. I also became aware of just how easy it was to be misled by this data as a result, even unintentionally – and this is on top of the typically expected misunderstanding or misinterpretation that can occur with any statistical data when given to the wider population.

In an event of national (and international) importance, clarity of communication is vital to prevent misinformation. I later tried to collate and document official definitions of Hospitalised cases [2], as well as having to learn the various definitions and nuances of other metrics. They represent best effort but may not be correct. The definitions were rarely easy to find or access, and I had to contact the respective health departments for some of them. This is not a step most of the general public would take to try to understand the meaning behind the numbers.

It wasn't until the latter half of 2021 when multiple states and territories (initially, NSW and Victoria, but later others) experienced significant and overlapping outbreaks, that comparisons of data across jurisdictions became increasingly important. This exposed many issues of the state-based approach to healthcare as it relates to statistics – each jurisdiction having differences in nearly every aspect of data collection and definitions.

This issue was further compounded by the media, as in the interests of simplicity (and time), journalists and reporters often relied on simplifications of data for public consumption. For example, a "Hospitalised case" was treated as or assumed to be equivalent across jurisdictions, leading to direct comparisons being made (e.g., "Jurisdiction A has twice as many reported, therefore it must be twice as bad there"). In reality, definitions for metrics differed wildly and often made such comparisons meaningless without a nuanced understanding and discussion of the data, which was rarely present.

While there are numerous examples of different definitions in use, two in particular are noteworthy because of the impact they have had on perceptions of how states have handled the pandemic (especially New South Wales and Victoria): Hospitalised cases, and Deaths.

Hospitalised COVID-19 cases

This metric has a wide variety of definitions, owing to the differences between hospital admissions vs beds occupied, inclusion of active vs inactive cases, and the respective definitions of an active case. NSW has, as a result, reported significantly higher hospitalisation metrics than other jurisdictions across the pandemic. Many reports have attributed this to a worse handling of the pandemic by NSW health, or to various other causes. But they are not actually comparable in any meaningful way, and do not support that conclusion.

COVID-19 deaths

While the ABS has a standardised methodology in reporting deaths from COVID-19 (or where COVID-19 was a contributing factor), the definitions used by jurisdictions are not standardised, in addition to the differences between when a death was reporting compared to when it occurred. New South Wales use the Registry of Births Deaths and Marriages (BDM) [3], resulting in a reasonably accurate accounting of deaths compared with the more accurate subsequent ABS data.

Victoria, however, continue to use a definition [4] that includes estimates based on recent infection, and their reported deaths due to COVID-19 in weekly state based reported are significantly over-represented and differ greatly from what ABS ultimately report. During many periods, this difference resulted in the appearance of a far higher death rate in Victoria compared with New South Wales and elsewhere, despite later ABS reports not indicating the same.

From late 2022, jurisdictions started to rapidly change their reporting – typically by scaling back the information collected and made available about COVID-19. While mostly a result of the changing nature of the pandemic and related legislation, these changes were numerous and difficult to track even for those used to handling this data. I have been recording changes to the official data state, territory and federal data sources for COVID-19, and what information they make available [5].

Typically done haphazardly, often only announced at the time of the change (or not at all), they resulted in significant effort on the parts of the various parties who maintained tracking sites or otherwise collected and performed analysis of this data. Covid Base AU have discontinued updates citing the challenge of finding data. The changes also introduced difficulties in explaining the data (e.g., trends), when the changes resulted in a discontinuation in the ability to compare new data with previous data, due to definition or criteria changes. The addition and later removal of cases detected via Rapid Antigen Test is one such example of a change to the metric that can falsely skew the apparent trend over time and prevent accurate comparisons. But other factors such as changes to test reporting requirements, test availability and various legislation also have an effect that is often overlooked.

The formation of a standardised, minimum set of reported metrics across all jurisdictions (along with a systematic approach to reviewing this) would have benefited the community greatly during this pandemic. The definitions should have been readily accessible; the formats, frequency, and availability of this data consistent across jurisdictions. Changes to these should have been both communicated in advance, and a history kept so that when looking back through past data, the changes are easily determined. Such changes should also be noted in official sources when visualising data if they materially change the interpretation of it.

---

[1] In the Media
https://github.com/dbRaevn/covid19/blob/main/pandemic/Media.md

[2] Official COVID-19 data definitions
https://github.com/dbRaevn/covid19/blob/main/pandemic/Definitions.md

[3] New South Wales COVID-19 surveillance report data sources and methodology
https://www.health.nsw.gov.au/Infectious/covid-19/Pages/surveillance.aspx

[4] Victorian COVID-19 surveillance report (definition of COVID-19 deaths metric found in appendix)
https://www.health.vic.gov.au/infectious-diseases/victorian-covid-19-surveillance-report

[5] Official COVID-19 data sources
https://github.com/dbRaevn/covid19/blob/main/pandemic/Datasources.md