Unlocking COVID-19 testing data: Best practices for public reporting

Our World in Data
2 June 2020

The importance of clear, accessible and timely data on testing

As emphasized by the WHO, extensive testing is a central countermeasure in the fight against COVID-19.[1]

The number of confirmed cases and deaths is what informs us about the development of the pandemic and guides our responses, but the confirmation of a case is based on a test.

Reliable data on testing is therefore necessary to assess the reliability of the data that informs us about the spread of the pandemic: the data on cases and deaths. The collection, dissemination and use of good data on testing is a key part of the global response to the pandemic.

Our World in Data is currently producing a [COVID-19 Testing Dataset](#)[2] by collecting official data on the number of tests performed, as published by different countries through a variety of official channels, including health ministry websites, press releases, and official social media accounts. This dataset is in high demand, and it is relied upon by many citizens, journalists, policymakers and health researchers around the world.[3]

Limitations with current reporting practices

Countries do not systematically report testing data to any international authority, or according to any international standard.

This lack of standardization of protocols and coordination leads to large differences in reporting practices. Some countries report their testing data very well, but many, either do not publish testing data at all, or do so in a way that limits proper use by others.

These differences in how countries report their testing data have real consequences. Our understanding of testing in different countries across the world, and hence our understanding of the pandemic itself, is less complete than it might be if reporting practices were improved.

---

[1] WHO Director-General's opening remarks at the media briefing on COVID-19 - 16 March 2020. See www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---16-march-2020

[2] ourworldindata.org/coronavirus-testing

[3] For a selected list of citations, see ourworldindata.org/coverage#coverage-in-2020

Towards better reporting practices

Based on our experience of collecting, presenting and interpreting testing data published by more than 80 countries, we highlight here a set of good practices for providers of official data on COVID-19 testing.

We outline two broad areas for improvement:

- Improving *access* to the available data;
- Improving the *documentation* of published data.

The recommendations outlined below are already followed by many health authorities in both high- and low-income countries. If those good practices can be followed by all countries, then citizens, policymakers and researchers will be in much better position to fully utilize the available testing data and put it to work in understanding, and ultimately combatting, the pandemic.

Best practices for data *access*

We suggest that national authorities adapt the way they report data according to the following guidelines:

1. **Make the data public:** up-to-date figures should always be released publicly, rather than sent to international organizations or websites such as *Our World in Data* privately (such as by email).

2. **Publish in a consistent location:** for the data to be noticed and collected promptly, the national authority should always release it in the same location (such as a COVID-19 dedicated website, a Ministry of Health press release section, etc.).

3. **Use a machine-readable format:** the data should be published using machine-readable formats such as CSV, JSON, XLSX, or ODS. When not possible, the second-best option is an HTML table. Countries should avoid publishing their data as text (with the relevant figure in the middle of a paragraph) or in PDF files.

4. **Make the data available at a stable URL:** data published behind changing URLs (URLs that include timestamps, changing random IDs, etc.) or single-use blob URLs mean that data collection cannot be done easily and automatically. Instead, the data should be accessible behind a stable and public URL (e.g. https://xxxxx.xx/data.csv)

5. **Update the full time series:** rather than a daily snapshot of the cumulative total number of tests done since the start of the outbreak, the data should be a full time series that is updated daily with a new row for the day.

6. **Follow typical guidelines for data publication:** the published data should follow the typical conventions of data storage, such as <u>ISO 8601</u> for dates, <u>snake case</u> for column names, <u>UTF-8 encoding</u>, etc.

7. **Provide detailed data:** headline aggregate figures should be complemented, where available, by breakdowns in terms of age group, region, or the testing process employed.

Belgium is a good example of a country that follows all of these guidelines:

➔ The Belgian institute for health, makes the data available <u>on a public website</u>.[4]
➔ The data has been published consistently on that page since the start of the outbreak.
➔ All variables are available in CSV and JSON formats.
➔ The files are accessible at stable and public URLs.
➔ Each file contains the full time series with all data points since the outbreak started.
➔ Dates and variable names are in a consistent and easy-to-process format.

Best practices for data *documentation*

Data on testing cannot be used well without an understanding of what published figures mean. A number, in the absence of a description, tells us little.

This is all the more important given differences across countries in the way testing data is collected, aggregated and presented. In different countries figures may relate to the number of people tested or the number of tests performed; they may include or exclude tests pending results; achieve different degrees of coverage in terms of reporting labs or regions; reflect different time periods; and so on.

This reduces the comparability of published figures across countries and over time, limiting what we are able to learn from the available testing data about the pandemic.

It is a problem that cannot be entirely solved. Some cross-country differences in the way testing data gets reported are a consequence of the different underlying approaches to testing that have been adopted. Moreover, they are somewhat inevitable in the absence of any international coordination or centralized collection of testing data.

But individual countries can do much to help improve the situation, simply by **describing the data they publish with clearly-written, detailed documentation.**

Where differences between countries' data are understood, any issues with comparability can be more accurately and reliably identified and assessed, greatly mitigating the problem.

---

[4] <u>epistat.wiv-isp.be/covid</u>

The following checklist provides the details needed for users of published testing data to properly interpret the data and compare across countries.

This information should be made available at an easy-to-find location alongside the data.

Checklist for COVID-19 testing data documentation

1. Which *units* are being counted?
It must be clear whether figures refer to the number of people tested, or the number of tests performed. Where possible, providing both figures, clearly labelled can be very useful for aiding cross-country comparisons.

2. Which testing technologies figures relate to?
It is essential that figures for different testing technologies are provided separately, and are clearly and unambiguously labelled (whether PCR, antibody, other types of antigen testing).

3. Are negative results included? Are pending results included?
As well as positive tests, data should include counts of negative test results. The units in each case (whether people tested or the number of tests) should be clear, and they should be consistent across positive and negative outcomes.

Figures for tests pending results should be reported separately. If, however, they are included in an aggregate figure, this should be made explicit. Many sources report the number of individuals who are 'suspected' or have been 'ruled out'. To be reliably included in test counts, it needs to be explicit whether such categories reflect the number of people who are awaiting test results or have tested negatively.

Test kits that have only been dispatched and not yet used should be reported separately from tests with or pending results.

4. Do the figures include all tests conducted in the country, or only some?
Figures reported by countries may only be partial if not all laboratories are reporting to the central authority. The scope of testing data should be made explicit by the source.

5. Are all regions and laboratories within a country submitting data on the same basis?
Answers to the questions above may vary from region to region. In order to assess the reliability of aggregate testing data, it needs to be clear if heterogeneous data is being summed together.

6. What period do the published figures refer to?
It should be clear whether figures relate to the date on which samples were taken, processed, or reported.

Cumulative counts of the total number of tests should also make clear the date from which the count begins.

Because the reporting of tests can take several days, for some countries figures for the last few days may not yet be complete. It needs to be made clear by the source if this may be the case.

7. Are there any issues that affect the comparability of the data over time?
To consider how testing figures are changing over time, it needs to be understood how any of the factors discussed above may have changed too.

8. What are the typical testing practices in the country?
Having a sense of how often and when individuals are tested, can help the users of these statistics understand how estimates of tests performed and individuals tested might relate to each other.

For instance, how many tests does a case investigation require? What are the eligibility criteria to be tested? Are health workers, or other specific groups, being routinely retested?

Our World in Data is a collaborative effort between researchers at the University of Oxford, based at the Oxford Martin Programme on Global Development, and the non-profit organization Global Change Data Lab.

Our data and writing on COVID-19 testing can be found at: ourworldindata.org/coronavirus-testing.

Further enquiries concerning the *Our World in Data COVID-19 Testing Dataset*, or good reporting practices, may be made to Joe Hasell, at joe@ourworldindata.org.