Making Sense of COVID-19 Data

Making Sense of COVID-19 Data

We are bombarded with data and information. The morning news reports the daily tally of confirmed COVID-19 cases and deaths, disaggregated by country, state, and city. Similarly, modelers produce widely disparate projections of COVID-19-related deaths in the United States that change daily. Ideally, this wealth of information should help inform governments, organizations, and individuals and lead to better decisions. However, it can be overwhelming and have the opposite effect.

Cutting through the clutter is critical. How we interpret information can have real-world, sometimes life or death, consequences. An article in The Colorado Sun on April 7, 2020 described the challenges the state government faces in preparing the health system to treat COVID-19 patients when three different models offer widely differing projections of the number of cases. It's a good time for us to remind ourselves how to be savvy consumers of data and information.

Guidelines for Interpreting Information

Which information can you trust? Unscrupulous actors often "lie with statistics," deliberately misrepresenting data to influence opinions or decisions. But, when filtered through mass media, even information from competent, well-intentioned individuals or organizations may unintentionally mislead. Information is analyzed data. Appraising how data are collected, analyzed, and presented can help you assess the "quality" of the information you consume and on which you base decisions.

Consider the Source

Do you trust the impartiality of the information source? Ask yourself whether the information provider has an ulterior motive to influence, or reinforce, your belief or behavior. A prejudiced information source may use suspect data. But even when the data come from a trusted source, the analysis or presentation may still produce misleading information.

This paper was prepared by MSI staff to contribute to the discussion and understanding of the important development challenges facing policymakers and practitioners

Assess the Data

Most consumers of information will not have the time, resources, or access to delve deeply into (sometimes proprietary) data to fully understand their characteristics or assess their quality. However, some potential issues with data are relatively easy to spot.

- All data are biased to some degree. Assess the likely magnitude and direction of bias in the data and how it may affect your interpretation of information derived from the data. Sampling plans and their implementation are common sources of bias. For instance, given what little we know about how individuals are selected for COVID-19 testing, can we reasonably infer anything about the incidence of the disease in the general population from reported data on the percentage of tests that are positive?
- Data are specific to a point in time and some—country-level socioeconomic data for instance—are often collected infrequently. Determine whether data are timely enough to provide relevant information. For relatively stable data, or data that can be reliably projected, older data may be fine. For rapidly evolving, unique situations, such as COVID-19, current data are crucial.
- The level of analysis is important. Assess whether the data are sufficiently granular to be relevant for decision-making. For instance, national-level data may not be appropriate for some decisions in countries with a great deal of regional variation.
- Data quality is multi-dimensional, variable, and difficult to assess, especially for secondary data. Reviewing data collection instruments and collection and analysis processes can provide some sense of likely quality. Questions that are ambiguous, require recall of hard-to-remember data, or ask about sensitive issues can compromise the validity of the data. Unusually large variability in a variable, or unexplained internal inconsistencies or patterns across variables may indicate errors in data collection or entry.

Understand the Analysis and Presentation

Carefully consider what you can reasonably infer from data and whether it is relevant to you. For example, the number of confirmed COVID-19 cases by location tells you something about the relative prevalence of the disease, if interpreted in the context of the detection regime. However, if you want to know the relative risk of exposure across locations, cases per capita may be more relevant.

> Confirm that claims of causality or attribution are based on robust and credible causal analysis.

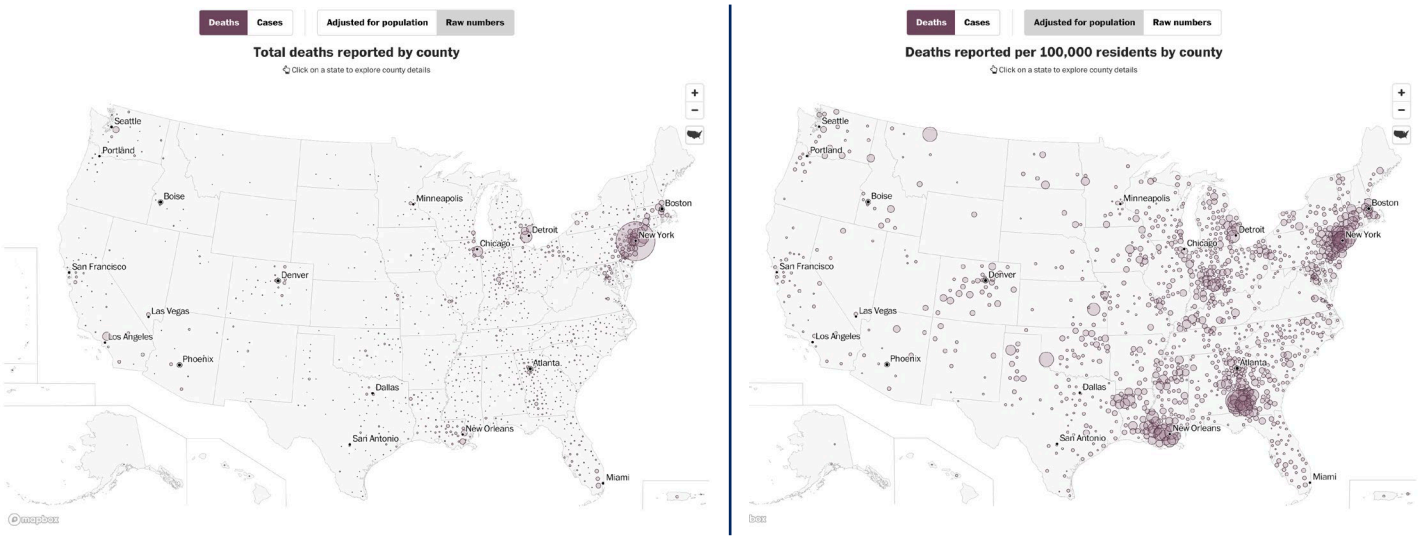
Remember, correlation does not imply causality. In the early stages of an outbreak of a new disease such as COVID-19, risk information often consists of observed correlations between a population parameter and a health outcome. For example, the Center for Disease Control and Prevention states that people 65 years of age

and older face a higher risk of severe illness from COVID-19. Presented with such information, it is easy to conclude that age is a meaningful factor in determining health outcomes. However, as scientists gather more data and refine their analyses, they may find that a factor correlated with age—pre-existing health conditions for instance—is the primary determinant of severe illness. Confirm that claims of causality or attribution are based on robust and credible causal analysis.

Unmeasured, or unmeasurable, data may also limit our understanding of a situation or relationship. For example, because of our inability to test everyone in the United States for COVID-19, data on confirmed cases or deaths may bear little resemblance to actual cases or deaths. Similarly, because of differences in testing regimes, data on confirmed cases and deaths are probably not even reliably comparable across states.

Maps, charts, and infographics are visually appealing, but they are designed to highlight specific patterns in data. If these visual products are not well-designed or clearly explained, they can be difficult to interpret. For example, a map of the United States showing the number of tests conducted for COVID-19 in each state might be useful for some decisions. Other decisions, however, may require that the data be presented in per capita terms or include a description of the criteria for selecting who to test *(see below)*. Examine visual products carefully to make sure you understand them and their implications.

Complex explanatory or predictive models can be particularly difficult to interpret. You may need to consult with specialists to assess the quality and reliability of such models.



COVID-19 deaths by county in raw numbers and adjusted for population. ***(Source: Washington Post)***