


2020

## The COVID-19 Pandemic and the Power of Numbers

Jessica Ancker

Weill Cornell Medical College, [jsa7002@med.cornell.edu](mailto:jsa7002@med.cornell.edu)

Follow this and additional works at: <https://scholarcommons.usf.edu/numeracy>

 Part of the [Disease Modeling Commons](#), [Epidemiology Commons](#), [Health and Medical Administration Commons](#), [International Public Health Commons](#), [Medical Education Commons](#), [Respiratory Tract Diseases Commons](#), and the [Virus Diseases Commons](#)

---

### Recommended Citation

Ancker, Jessica. "The COVID-19 Pandemic and the Power of Numbers." *Numeracy* 13, Iss. 2 (2020): Article 2. DOI: <https://doi.org/10.5038/1936-4660.13.2.1358>

Authors retain copyright of their material under a [Creative Commons Non-Commercial Attribution 4.0 License](#).

---

## The COVID-19 Pandemic and the Power of Numbers

### Abstract

The COVID-19 pandemic has produced a deluge of news coverage of quantitative concepts. In this viewpoint, we provide examples of effective and poor quantitative communication by the professional news media as well as social media communicators. Effective examples include a number of online animations and engaging interactive simulations. Examples of poor quantitative communication include the widespread reporting of raw numbers rather than rates, failing to address uncertainty, not providing sufficient context for numbers, and not discussing the implications of false negative and false positive diagnostic test results. Educators can draw from this body of news to develop compelling quantitative literacy lessons but can also use informal means to disseminate high-quality quantitative information.

### Keywords

public health, epidemic, quantitative literacy

### Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/)

### Cover Page Footnote

Jessica Ancker is an associate professor of health informatics at Weill Cornell Medical School. Her research focuses on the use of health information technology to improve medical decisions. She teaches research methods and introductory statistical concepts to graduate students in health informatics and medical students at Weill Cornell. In addition, she gives an annual guest lecture series on statistics for journalists at the Columbia Journalism School. Dr. Ancker holds an MPH in biostatistics and a doctorate in biomedical informatics, both from Columbia University.

## Introduction

Pandemic news is crammed with numbers: the basic reproductive rate for the virus; the exponential curve; the case-fatality rate; the state, national, and international case count; the hospital bed count; the count of intensive care units; the safe distance for social distancing, and the proportion of people doing it.

Certainly no event in recent history has spurred such a deluge of media coverage and social media discussion about quantitative concepts. In a matter of weeks, politicians went from citing the case count to citing sophisticated mathematical models predicting hospital capacity and fatality rates. Magazines published so-called “explainers” on the difference between linear and exponential increase, and “doubling time” became a common phrase. A new generation of journalists created eye-catching interactive graphics illustrating models of disease spread as a function of duration of social distancing, and animated maps of cell phone data showing volume of travel from disease hotspots. Social media allowed professional and amateur data scientists, armed with easy access to data and software, to share memorable graphs and data-rich memes.

This is the pandemic represented in the public imagination by, of all things, a meme containing a pair of bell-shaped curves (see Appendix for links to all illustrations and animations). The first, tall and thin, represents the disease caseload if we do nothing. The longer and lower arch, limboing beneath a line representing hospital capacity, popularized the phrase “flatten the curve” (Godoy 2020).

During this unprecedented time, we have the opportunity to examine some ways in which media coverage and social media interactions have advanced quantitative literacy, and some ways in which they have not.

## Media Successes

One amazing development has been the extremely effective use of animation and interactivity to allow anyone with a good Internet connection to view and even explore data to better understand complex concepts.

One example is the innovative simulations by Harry Stevens at the *Washington Post*, which provide a bird’s-eye view of groups of people passing the virus to each other under different assumptions: uncontrolled conditions; attempted quarantine (in which a “wall” segregates a vulnerable group); moderate social distancing observed by 25% of the population; and extensive social distancing observed by 75% of the population (Stevens 2020). There are oversimplifications, but they are clearly stated (for example, in the simulations, every individual who comes in contact with an infected person gets the disease). These graphics deliver not only insight into disease spread, but also an emotional impact from seeing individual people affected.

Another example is a more powerful multivariate interactive simulation developed by Stanford researcher Erin Mordecai and her team to show the impact of social distancing under a variety of different parameter assumptions (Mordecai 2020). This site gives the motivated reader the ability to explore how different assumptions shape and shift the curves. The researchers' sobering conclusion becomes more credible the more we explore the simulation: no matter what we alter, we almost always get a disease resurgence later in 2020.

Even without the power of simulation modeling, creative animated graphics can provide new insight into both scale and change over time.

For example, a gif of unemployment claims on April 2 by Twitter user How Things Work (@ThingsWork) provides a visceral understanding of the economic impact of the pandemic (2020). The designer cleverly adjusted the y-axis as time progresses, which means that the 20th-century trends are placed in appropriate context. Then the graph explodes vertically so that the massive March 2020 spike is felt as a gut punch.

To show how quickly COVID-19 became the leading daily cause of death, another regularly updated animation shows the ranking changing rapidly in recent weeks (Danilychev 2020). Similarly, these animated rankings by biostatistician Ivy Chen show how states and countries jockeyed for their unenviable positions as leaders in COVID-19 case counts (Chen 2020).

Without animation, these explanatory graphics would have required the reader to do considerably more cognitive work. Readers would have had to compare a sequence of side-by-side graphs, or disentangle several curves superimposed on the same graph (as in this more traditional graph from *The Lancet* (Anderson et al. 2020)). Alternately, they might have had to look at graphics showing deltas rather than data values. Because abstract graphs are much harder for quantitative novices to understand, it's unlikely they would have been as effective in conveying the take-home message.

## **Media Missteps**

### ***Emphasizing Raw Numbers***

First, the news and policy announcements tend to focus on raw numbers, rather than data standardized by unit population.

This clever animated map from the April 6 *New York Times* showing numbers of deaths rising over time was published on the occasion of the United States reporting a total of 10,000 deaths (Gamio and Yourish 2020). But none of the numbers are standardized. Numbers are higher in metropolitan areas, but without the population size, it's completely unclear how much of the difference between (say) Chicago's toll and Boston's is a function of their relative sizes. To their credit, the paper also posted per-capita data; a bit of searching through the less visually

arresting line graph lower on the same page reveals that Boston and Chicago actually have exactly the same per-capita death rate (3 per 100,000).

The *Washington Post* on April 7 reported the sad news that grocery workers were succumbing to coronavirus because they were continuing to work, often without personal protection, during the epidemic (Bhatterai 2020). As tragic as this is, the numerator isn't terribly meaningful without the denominator. What proportion of grocery workers are falling victim, and how does it compare to the proportion in other professions? The argument about the importance of this number would be even stronger if the proportion could be compared with the proportion for other known high-risk professions, such as medicine and public safety. Widespread coverage of 41 deaths among New York City's Metropolitan Transit Authority workers similarly omitted any mention of how large the MTA workforce is (Goldbaum 2020).

A peculiar entry in the April 6 *USA Today* is a final example of how misleading raw numbers can be, especially in the absence of denominators. The article pointed out that the US death toll had surpassed the combined total from six previous American wars (Shannon 2020). The author had to cherry-pick a list of wars to support the thesis. The list includes the American Revolution—which took place with a total contemporary population of less than 3 million, less than 1/100th the current US population.

### ***Failing to Address Uncertainty***

The second area in which we need to see improvement is in the way the media handle uncertainty. Everything about this epidemic is tinged with uncertainty, but nowhere more obviously than in the number of COVID cases. The case count is dependent on the availability of tests, and that means it is unreliable, especially in the United States, given our late start to testing, our very low rate of testing, and our practice of reserving tests for the sickest patients. Most people with symptoms, especially in the hardest-hit regions, are being told to stay home without being tested, so we may never know the true case count. And because the case count serves as the denominator of the hospitalization rate and the case-fatality rate, these figures are also unreliable.

Nevertheless, case counts are reported daily by news organizations, in part because they are readily available (Johns Hopkins Coronavirus Resource Center 2020). On March 26, many news organizations reported that the US had surpassed China in total cases. Most of this coverage made no mention of the testing shortage in the United States (or the four-fold difference in the countries' populations). One example, in the *Los Angeles Times*, also states that locally, “. . . 253 people were at some point hospitalized, which amounts to about 21% of all positive cases . . . ,” uncritically using the case count as the denominator for the hospitalization rate calculation (Lin II 2020). The true hospitalization rate would require the total

number of cases as its denominator, which unfortunately is not known. The governor of Minnesota recently attributed a slowing in the rate of new infections to social distancing, but the same news article also noted that the state health lab was conducting testing primarily in long-term care facilities, raising important questions about whether the apparent slowing was an artifact of the testing approach (Olson 2020).

However, in addition to these examples of misleading denominators, there are good examples as well, such as the maps and graphics displaying results of *The New York Times*' impressive case-tracking effort (2020). In these, death rates are reported both in raw numbers and as per capita population rather than as a proportion of those infected.

A National Public Radio graphic does an excellent job of portraying a different sort of uncertainty, the confidence interval or margin of error, by superimposing an orange cloud over the projected curve of epidemic deaths in each state (McMinn 2020). This could be further improved with a better text explanation of the cloud, currently described simply as the "estimate range." This could be an excellent opportunity to explain to a motivated readership that researchers always have uncertainty when making an estimate on the basis of insufficient data.

### ***Failing to Put Numbers in Context***

A very simple recommendation for improvement is for journalists to do a better job of putting unfamiliar numbers in context. If I am a reader encountering a statistic that I've never heard of before, I need help understanding why it is important. Is the number high or low? Should I be alarmed by this number, or reassured by it? Can I trust it? For workshops on numbers and statistics that I teach for journalists, I developed a checklist of questions journalists can use to select five types of context that will help their readers understand novel numbers. The last time I taught it to science journalists (in March 2020), we worked together to apply it to the COVID-19 "basic reproductive number." The five-question checklist resulted in these answers:

1. *What does the number mean in English and why is it important?* The basic reproductive rate tells us how contagious the disease is. A basic reproductive rate of 2.3 would mean that each infected person passes it to an average of 2 to 3 others.
2. *What's the possible range of the number?* A non-transmissible disease has a basic reproductive rate of 0. The basic reproductive rate of measles, one of the most contagious diseases, is about 18.
3. *Are there important categories or thresholds to interpret it?* A basic reproductive rate less than 1 indicates the disease will die out spontaneously.
4. *What comparison values might help the reader understand the importance of the number?* The basic reproductive rate for seasonal influenza is usually between 1 and 2, indicating that COVID-19 is much more contagious than ordinary flu.
5. *Is there uncertainty about the number?* Yes. We have inadequate data, especially about how many people actually have the disease. Also, the data we have is rapidly changing.

Although it's the best we can do right now, there's a tremendous amount of uncertainty about this number, and we can expect it to change.

The news media has been uneven in terms of putting unfamiliar coronavirus numbers in context. For example, a glossary published March 18 defines “basic reproductive number” and gives the important information about the threshold of 1, but gives no other context to show why this number is so concerning (Gross and Padilla 2020). By contrast, a CTV article from March 3 contains all five of the contextual elements (Flanagan 2020).

### ***Failing to Address the Weaknesses of Diagnostic Tests***

To date, news coverage of COVID-19 testing has promoted the idea that the tests will provide the right answer, and that our societal problem is a lack of widespread testing.

But in fact, the performance of our current tests for the virus is not optimal, as respected health researcher Harlan Krumholz points out (2020). One type of test that would be extremely helpful is an antibody test, which could be performed to determine whether an individual has previously been exposed to the virus and might now be immune. However, tests in development still produce not only false negatives (failing to identify people with antibodies), but also false positives (erroneous positive results among people who do not have COVID-19 antibodies).

What an opportunity this creates to promote understanding of Bayes' Theorem (Boersma and Willard 2008)! Suddenly, this theorem is central to the problem of whether we could use an antibody test to distinguish immune from non-immune Americans at scale.

If we were to develop a test with 90% sensitivity, that would mean for every 100 people with antibodies, our test would correctly capture 90 of them. And if our test had 86% specificity, then that would mean for every 100 people without antibodies, our test would correctly identify 86 of them, but it would wrongly classify the remaining 14 people as antibody-positive.

The difficulty arises when we scale up our testing program to the entire population. Imagine what would happen if we applied this test at a time when about 12% of the US population (about 38.4 million people) had been exposed to coronavirus and were antibody-positive. When we gave the test to the 38.4 million, we would correctly diagnose 90% of them, producing almost 34.6 million positive test results.

Unfortunately, when we administered the test to the remaining 281.6 million antibody-negative people, we would be wrong 14% of the time, producing another 39.4 million positive test results, all false positives.

This would create an enormous problem. A lot of people who were not immune would think that they were—in fact, more than half of the 74 million people who

received positive test results and thought they were immune would be wrong. Clearly, this test would not be suitable for widespread screening.

And yet if we applied exactly the same test (90% sensitivity, 86% specificity) to a population in which 95% of people had antibodies, we'd produce 273.6 million true positives and only 2.2 million false positives. In this situation, more than 99% of all the positive test results are true positives, greatly increasing our confidence in any individual positive result.

It's not just the antibody test that is important here. Krumholz points out that the actual COVID-19 test itself has low sensitivity, meaning that there are almost certainly a lot of false negative results (erroneous negative results among people who actually have the virus) (2020). Because false negatives occur only among patients who truly are positive for the disease, they become common when the test is administered to a group with very high prevalence of disease, such as symptomatic patients in disease hotspots. Krumholz points out that this chance of a false negative means a person who has serious symptoms of COVID-19 should be treated for it even if the test comes back negative (2020).

I have yet to see a quantitatively literate discussion of why this means it would be a mistake to rush a poorly performing antibody test to market, or of how disease prevalence in the population (also known as "prior probability" of disease) affects how we interpret the results of a medical test, even a test with good performance.

## Next Steps

As educators, we can and should leverage this crisis into a teaching opportunity for those reading the news. We should share reliable data, insightful analyses, and good graphics. We should debunk problematic interpretations of the data. We should exploit the public hunger for good information by, as others have argued, developing teaching opportunities for "citizen statistics," covering "the kinds of numbers that describe and delineate our personal and public lives" (Hacker 2012). We can do this by updating the quantitative literacy courses we already teach with the rich set of examples from the current pandemic, and by providing exercises to help our students identify and see through sloppy use of numbers in the news. We should also exploit more informal opportunities through our own social networks. At least once a day, I use social media to answer questions or try to correct misperceptions. After asking me what exponential increase really meant, one friend manually created a list projecting day-by-day case counts under the assumption of a 3-day doubling time. Each day, he checks it against the published numbers, and he correctly projected the day when the nation would hit 100,000 cases. One small step for "citizen statistics."

These small steps, combined with the current public hunger for good information, might help raise support for bigger changes. Each person who



estimates case counts on the basis of the exponential curve, and then observes how the real world matches the projection, gets a better understanding of how the world works. That understanding could possibly engender some trust in the data or even some confidence in the policy decisions based upon it.

## References

- Anderson, R.M., H. Heesterbeek, D. Klinkenberg, T.D. Hollingsworth. 2020. “How Will Country-based Mitigation Measures Influence the Course of the COVID-19 Epidemic?” *Lancet*, 395(10228):931–4.
- Bhatterai, A. 2020. “Grocery Workers Are Beginning to Die of Coronavirus: At Least Four People—Who Had Worked at Walmart, Trader Joe’s and Giant—Have Died from COVID-19 in Recent Days.” *The Washington Post*, April 6. <https://www.washingtonpost.com/business/2020/04/06/supermarket-workers-deaths-coronavirus/>
- Boersma, S., and T. Willard. 2008. “False Positives and Referral Bias: Content for a Quantitative Literacy Course.” *Numeracy*, 1(2).
- Chen, I. 2020. “NYC’s COVID-19 Crisis.” <https://www.biostatistically.com/nyc-covid-19-crisis/>
- Danilychev, M. 2020. “COVID-19 Daily Deaths vs. Top 15 Causes of Death (Average/Day) in the US.” <https://public.flourish.studio/visualisation/1727839/>
- Flanagan, R. 2020. “How Contagious Is COVID-19 Compared to Other Viral Diseases?” *ctvnewsca*, March 3. <https://www.ctvnews.ca/health/coronavirus/how-contagious-is-covid-19-compared-to-other-viral-diseases-1.4836734>
- Gamio, L., and K. Yourish. 2020. “See How the Coronavirus Death Toll Grew across the US.” *The New York Times*, April 6. <https://www.nytimes.com/interactive/2020/04/06/us/coronavirus-deaths-united-states.html>
- Godoy, M. 2020. “Flattening a Pandemic’s Curve: Why Staying at Home Now Can Save Lives.” National Public Radio, March 13. <https://www.npr.org/sections/health-shots/2020/03/13/815502262/flattening-a-pandemics-curve-why-staying-home-now-can-save-lives>.
- Goldbaum, C. 2020. “41 Transit Workers Dead: Crisis Takes Staggering Toll on Subways.” *The New York Times*, April 8. <https://www.nytimes.com/2020/04/08/nyregion/coronavirus-nyc-mta-subway.html>.
- Gross, J.P., and M. Padilla. 2020. “From Flattening the Curve to Pandemic: A Coronavirus Glossary.” *The New York Times*, March 18. <https://www.nytimes.com/2020/03/18/us/coronavirus-terms-glossary.html>

- Hacker, A. 2012. "Is Algebra Necessary?" *The New York Times*, July 28.  
<https://www.nytimes.com/2012/07/29/opinion/sunday/is-algebra-necessary.html>.
- How Things Work. 2020. Twitter post. April 3, 2:59 p.m.  
<https://twitter.com/ThingsWork/status/1246150462073536512?s=20>
- Johns Hopkins Coronavirus Resource Center. 2020. "COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University." <https://coronavirus.jhu.edu/map.html>
- Krumholz, H.M. 2020. "If You Have Coronavirus Symptoms, Assume You Have the Illness, Even If You Test Negative." *The New York Times*, April 1.  
<https://www.nytimes.com/2020/04/01/well/live/coronavirus-symptoms-tests-false-negative.html>
- Lin II, R.-G. 2020. "US Surpasses China, Italy in Coronavirus Cases; California Sees Spike as Well." *Los Angeles Times*, March 26.  
<https://www.latimes.com/california/story/2020-03-26/u-s-surpasses-china-italy-in-coronavirus-cases-california-sees-spike-as-well>.
- McMinn, S. 2020. "Coronavirus State-by-State Projections: When Will Each State Peak?" National Public Radio, April 7.  
<https://www.npr.org/sections/health-shots/2020/04/07/825479416/new-yorks-coronavirus-deaths-may-level-off-soon-when-might-your-state-s-peak>
- Mordecai, E. 2020. "Potential Long-Term Intervention Strategies for COVID-19." <https://covid-measures.github.io/>
- The New York Times*. 2020. "Coronavirus Map: Tracking the Global Outbreak." <https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html>
- Olson, J. 2020. "Walz to Tweak Stay-at-Home Restrictions as Minnesota Battles COVID-19." *Minneapolis StarTribune*, April 8.  
<https://m.startribune.com/minnesota-covid-19-cases-over-1000-with-34-deaths-veterans-home/569440222/>
- Shannon, J. 2020. "The US Has Surpassed 10,000 Deaths from Coronavirus. That's More than the Total Battlefield Fatalities from Six Combined Wars." *USA Today*, April 6.  
<https://www.usatoday.com/story/news/health/2020/04/06/coronavirus-death-toll-us-reaches-10000-six-wars-combined/2949285001/>
- Stevens, H. 2020. "Why Outbreaks Like Coronavirus Spread Exponentially, and How to 'Flatten the Curve'." *The Washington Post*, March 14.  
<https://www.washingtonpost.com/graphics/2020/world/corona-simulator/>

## Appendix

**Table 1**  
**Graphics and Animations Illustrating the Impact of the COVID-19 Pandemic**

Title	URL	Description
Flattening a pandemic's curve	<a href="https://www.npr.org/sections/health-shots/2020/03/13/815502262/flattening-a-pandemics-curve-why-staying-home-now-can-save-lives">https://www.npr.org/sections/health-shots/2020/03/13/815502262/flattening-a-pandemics-curve-why-staying-home-now-can-save-lives</a>	Two overlapping bell-shaped curves represent case count with and without social distancing
Why outbreaks like coronavirus spread exponentially, and how to "flatten the curve"	<a href="https://www.washingtonpost.com/graphics/2020/world/coronavirus-simulator/">https://www.washingtonpost.com/graphics/2020/world/coronavirus-simulator/</a>	Interactive simulations of how contact between infected and uninfected people spreads the disease
Potential Long-Term Intervention Strategies for COVID-19	<a href="https://covid-measures.github.io/">https://covid-measures.github.io/</a>	Powerful interactive simulation modeling the effect of different public health measures
US unemployment claims in their historical context	<a href="https://twitter.com/ThingsWork/status/1246150462073536512?s=20">https://twitter.com/ThingsWork/status/1246150462073536512?s=20</a>	Time trend of unemployment claims over the past century
COVID-19 daily deaths vs. top 15 causes of death (average/day) in the US	<a href="https://public.flourish.studio/visualisation/1727839/">https://public.flourish.studio/visualisation/1727839/</a>	Animation showing how COVID became a leading cause of mortality
NYC's COVID-19 crisis	<a href="https://www.biostatistically.com/nyc-covid-19-crisis/">https://www.biostatistically.com/nyc-covid-19-crisis/</a>	Animation of NYC statistics
How coronavirus's death toll grew across the US	<a href="https://www.nytimes.com/interactive/2020/04/06/us/coronavirus-deaths-united-states.html">https://www.nytimes.com/interactive/2020/04/06/us/coronavirus-deaths-united-states.html</a>	US map with hotspots represented as peaks
COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University	<a href="https://coronavirus.jhu.edu/map.html">https://coronavirus.jhu.edu/map.html</a>	Respected and regularly updated source of COVID-19 data
Coronavirus map: Tracking the global outbreak	<a href="https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html">https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html</a>	US tracking map showing mortality per capita
Coronavirus state-by-state projections: When will each state peak?	<a href="https://www.npr.org/sections/health-shots/2020/04/07/825479416/new-yorks-coronavirus-deaths-may-level-off-soon-when-might-your-state-s-peak">https://www.npr.org/sections/health-shots/2020/04/07/825479416/new-yorks-coronavirus-deaths-may-level-off-soon-when-might-your-state-s-peak</a>	State-specific curves projecting disease peaks