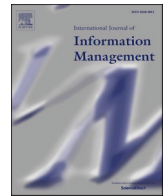




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Opinion Paper

## The challenges of data usage for the United States' COVID-19 response



S.E. Galaitsi\*, Jeffrey C. Cegan, Kaitlin Volk, Matthew Joyner, Benjamin D. Trump, Igor Linkov

U.S. Army Corps of Engineers, Vicksburg, MS, United States

## ARTICLE INFO

## Keywords:

Coronavirus  
Decision making  
Public health data

## ABSTRACT

During the coronavirus pandemic, policy makers need to interpret available public health data to make decisions affecting public health. However, the United States' coronavirus response faced data gaps, inadequate and inconsistent definitions of data across different governmental jurisdictions, ambiguous timing in reporting, problems in accessing data, and changing interpretations from scientific institutions. These present numerous problems for the decision makers relying on this information. This paper documents some of the data pitfalls in coronavirus public health data reporting, as identified by the authors in the course of supporting data management for New England's coronavirus response. We provide recommendations for individuals to collect data more effectively during emergency situations such as a COVID-19 surge, as well as recommendations for institutions to provide more meaningful data for various users to access. Through this, we hope to motivate action to avoid data pitfalls during public health responses in the future.

## 1. Introduction

Challenges to data collection and dissemination for epidemiology are not new. In 1995, *Science* examined factors like systematic errors in studies (e.g., confounding factors and biases), assessing exposure to risk factors, biases in the recorded data, an overemphasis on the statistical significance of results, and ways in which the media intentionally or unintentionally misrepresents the results of epidemiological studies (Taubes and Mann, 1995). These concerns have been reiterated and updated for the modern technological world in recent years to include the need for data to be available in common open formats, for explicit and clear time intervals and case definitions, and for a streamlined reporting process to reduce time lags (Fairchild et al., 2018). Nonetheless, decision makers have faced data challenges during the COVID-19 pandemic, which has brought the challenges into public discussion and scrutiny as government agencies, academic institutions, the media, and the public have stumbled into one pitfall after another.

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, requires policy makers to apply epidemiological information to guide their decisions on behalf of U.S. public health. These decisions include:

- 1) Is it safe for children to go to school?
- 2) How many people can safely gather in a bar, restaurant, or religious institution?

- 3) Should people wear masks on public transit, inside grocery stores, and outside?
- 4) How much personal protective equipment (PPE) should be distributed and when?
- 5) Where should limited medical resources, such as ventilators or therapeutics, be deployed?

To make informed public health decisions, policy makers need to understand characteristics of the disease and population susceptibility to it, including prevalence and mortality rates, as well as the effectiveness of protective measures. Thus, policy makers require that meaningful data be collected and made accessible (Fahey and Hino, 2020). However, the coronavirus pandemic has revealed challenges surrounding data collection and access in the United States, with ramifications ranging from individual behavioral complacency to insufficient institutional material resource access to prevent unnecessary deaths. Furthermore, ambiguities around data have encouraged divergent interpretations that may have confused public health messaging. The WHO Pandemic Influenza Preparedness and Response document emphasizes the importance for countries to "maintain trust across all agencies and organizations and with the public through a commitment to transparency and credible actions" (World Health Organization, 2009). During a pandemic, risk communication and messaging from critical institutions is essential to the public response; the crisis

\* Corresponding author.

E-mail addresses: [stephanie.e.galaitsi@usace.army.mil](mailto:stephanie.e.galaitsi@usace.army.mil) (S.E. Galaitsi), [jeffrey.c.cean@usace.army.mil](mailto:jeffrey.c.cean@usace.army.mil) (J.C. Cegan), [kaitlin.m.volk@usace.army.mil](mailto:kaitlin.m.volk@usace.army.mil) (K. Volk), [matthew.d.joyner@usace.army.mil](mailto:matthew.d.joyner@usace.army.mil) (M. Joyner), [benjamin.d.trump@usace.army.mil](mailto:benjamin.d.trump@usace.army.mil) (B.D. Trump), [igor.linkov@usace.army.mil](mailto:igor.linkov@usace.army.mil) (I. Linkov).

<https://doi.org/10.1016/j.ijinfomgt.2021.102352>

Received 9 November 2020; Received in revised form 29 March 2021; Accepted 29 March 2021

Available online 2 April 2021

0268-4012/© 2021 Elsevier Ltd. All rights reserved.

management agency that creates established, authenticated and qualified mental models can make better decisions and engender public trust as a source of relevant crisis information and advice for public action (Bunker, 2020). Conversely, confusion in data presentation and interpretation can undermine public trust of the information presented and the agencies presenting it.

Good data management practice should critically evaluate the data being used and its limitations. Many decisions are made, models constructed, and simulations run based on available data, potentially amplifying misinformation stemming from data flaws. “Garbage in, garbage out” is a common adage in fields reliant on large datasets. So too do we say “perfection is the enemy of the good.” Conditions of high uncertainty require balancing these conflicting concepts in order to enable taking timely action. Pitfalls can be avoided by exploring the metadata, data cleaning, and acknowledging the limitations of specific datasets and how those limitations may affect the overall results.

The coronavirus pandemic represents an important moment of reckoning for scientists and policy makers to fully view the impacts of long-standing unresolved issues in public health and epidemiology. These issues compromised the speed and efficiency of policy responses that aimed to prevent unnecessary deaths. The pandemic has affected all research and its impacts will carry into the future (Barnes, 2020). Disruption of this scale offers a valuable opportunity to examine our existing information systems (Doyle and Conboy, 2020), and what it has revealed should both galvanize and empower us to do better. This paper catalogues various limitations of data availability, aggregation, and interpretation during the COVID-19 pandemic and provides recommendations to improve data management to better ensure relevance for decision makers. Section 2 explores the availability of reliable and clear data on COVID-19 outcomes as well as data characterizing the resources needed to ensure the safe and effective functioning of essential services. Section 3 examines aggregation and interpretation of published data. We examine implications for research in Section 4 and provide recommendations in Section 5.

## 2. Data availability

In this paper, we use the word “data” to refer to quantitative numerical values that support models and build metrics to provide a more complete picture of the COVID-19 situation in the United States. While data may also be qualitative, this is typically not the data we have been working with during the COVID-19 response. The importance of accurate, timely, and comparable data has been clear throughout the pandemic (Dwivedi et al., 2020), but many pitfalls still exist.

### 2.1. Defining data to prevent false comparisons

A critical use for data involves its aggregation across time and space to understand where outbreaks are and their trajectories. However, disparities in data definitions may cause aggregating organizations to effectively compare apples and oranges, and therein generate data that is somewhat meaningless or misleading. This section explores the pitfalls of inadequate or inconsistent data definitions and the consequences of trying to use that data to characterize temporal or spatial characteristics of COVID-19 outbreaks.

#### 2.1.1. Inadequate definitions

Much of data comes with qualification of nuances, but these are not always presented to data users. Positive COVID-19 case counts, for example, are typically represented by a single numerical value, with no indication of the uncertainty inherent in that value. Tests for COVID-19 are returned to doctors and patients from the testing facility as simply positive (the virus was detected) or negative (the virus was not detected). However, testing facilities that are utilizing polymerase chain reaction (PCR) tests have access to another number that would help clinicians and public health officials better determine active COVID-19

cases: the number of cycles run before the virus was identified.

During a PCR test in which COVID-19 is present, a genetic sequence specific to the coronavirus is replicated. Each cycle increases the amount of DNA present until the concentration is high enough for the PCR machine to detect. A positive result with a low number of cycles would indicate someone with high viral load, likely making them a high risk for transmission. In contrast, a positive result obtained only after a high number of cycles would indicate someone with a low amount of virus in their body, meaning they are in the early or late stages of their infection (Meyerowitz et al., 2020). Running a greater number of cycles increases the likelihood of detecting the virus in someone who is no longer contagious but still has residual virus in their body. Some experts say a reasonable cycle cut-off is 30–35. This would indeed miss some cases at the start of infection, but would not mischaracterize many more cases where the infection has passed. Yet many tests for COVID-19 are run up to 40 cycles (Mandavilli, 2020). Thus, treating all positive results equally means that people who are unable to spread the virus are being lumped in with those who can, blurring the meaning of a positive case.

Adding to this confusion, people who test positive for COVID-19 without symptoms are often collectively referred to as “asymptomatic” even though many of those people will eventually develop symptoms, and thus would be more accurately classified as “pre-symptomatic” (CDC, 2020a). Failure to differentiate between the two categories of cases can artificially inflate the estimated rate of asymptomatic carriers. This was a significant pitfall of one of the first larger studies of COVID-19 carriers in Massachusetts nursing homes (Goldberg, 2020).

Similarly, controversy has erupted surrounding what constitutes a death by COVID-19, with conservative media accusing states of over-counting COVID-19 deaths even as experts say under-counting of deaths is more likely (Walker et al., 2020) given that an increased number of deaths have been observed since March when compared to this timeframe in past years (Lu, 2020). The Netherlands switched to using this comparison between weekly deaths during the pandemic to weekly deaths in previous years to estimate deaths from COVID-19 after criticism that their lack of testing capacity was leading to an undercount (Janssen and van der Voort, 2020). Only 6 % of COVID-19 deaths in the US list COVID-19 as the sole cause (CDC, 2020b), whereas other causes of death may list co-morbidities in addition to COVID-19. However, Justin Lesser, infectious disease specialist at the Bloomberg School of Public Health, explains that the presence of co-morbidities does not mean patients did not die of COVID-19. He says that “COVID may have caused [the co-morbidities] or worked synergistically to kill them” (Pearce, 2020). Agencies within the same state may even have different definitions of what constitutes a COVID-19 death and should be counted in their state total. In Florida, for example, the Medical Examiners—who are responsible for certifying every COVID-19 death—and the health department separately track deaths. Their death counts do not match because the health department counts individuals who tested positive for COVID-19 and subsequently died (even if the death was caused by an unrelated accident) while the Medical Examiners count deaths that they have directly attributed to COVID-19 (McGrory et al., 2020). Additionally, the health department only counts deaths of Florida residents while the Medical Examiners count deaths of anyone who died in Florida (McGrory et al., 2020). Inconsistency in counting deaths during or after a disaster is not a new problem (Miceli, 2019). In the case of COVID-19, not having a uniform standard for counting deaths injects a substantial amount of uncertainty when trying to understand the state of affairs within a given locality or when comparing two localities against one another.

Field experts can use specific terminology when presenting raw data to assist other experts from the same field in understanding key differences and even uncertainties. However, this field-specific terminology may lack distinctions for lay people or government officials trying to draw their own conclusions. Raw data from different sources and collected using different methods can produce divergent conclusions that may be tied to definitions used in the studies. Column titles in

spreadsheets matter, but may be vague or confusing, and metadata explaining assumptions or definitions is often absent. At a time when data needs to move from research to policy quickly, these ambiguities can waste time, prevent interpretations, and perpetuate misunderstandings and misreporting.

This can extend to data characterization for experimental analyses. For example, two studies (Colaneri, Seminari, Novati et al., 2020; Colaneri, Seminari, Piralla et al., 2020) swabbed surfaces for SARS-CoV-2 RNA, then attempted to culture (i.e. grow “live” virus in a controlled environment) virus from the positive swabs. Their attempts were unsuccessful, supporting the conclusion of Mondelli et al. (2020) that SARS-CoV-2 is unlikely to be transmitted by surface contact in real world conditions, provided that standard cleaning procedures are followed. However, another study on “infectious SARS-CoV-2” survival found that infectious virus could survive for weeks on some surfaces at ambient temperature (68 °F) in the absence of sunlight (Riddell et al., 2020). The differences among these four studies may lie in the definition of “real world” conditions, “infectious SARS-CoV-2” or “survive.” Certainly, the laboratory methodologies may have also mattered in generating the final results, but for the lay reader, the studies seem to be claiming contradictory data findings despite examining the same type of data. The terms used to describe COVID-19 data should be clearly described so non-expert readers can understand the nuances of various data studies.

### 2.1.2. Inconsistent definitions across spatial dimensions

Different interpretations of data meaning would be minimized if a centralized authority provided precise definitions that specified the exact data types to be reported under various titles. For example, for the purpose of hospital reporting to the federal government, inpatient hospital beds are defined as staffed inpatient beds including ICU beds and all overflow and surge/expansion beds used for inpatients. However, in other respects the federal government has largely left states to each develop their own data management systems. Regional differences have emerged, for example, early in the pandemic, the CDC as well as Texas, Pennsylvania, Georgia, Vermont, and Virginia, reported positive viral tests and antibody tests in the same metric (COVID-19 cases), leaving no way for epidemiologists to differentiate between the number of people with current infections and past infections (Madrigal and Meyer, 2020). This makes it difficult to estimate the size of any current outbreak or to compare one state’s material needs to another’s. This error was corrected in late May 2020.

In regards to positive testing, an employee of the Florida Department of Health accused the state of artificially inflating the denominator by counting negative tests from the same person (i.e., total number of tests conducted). The data presented on the Florida website was not accurately labeled to reflect this practice (Wamsley, 2020). Other states are or may also be reporting positive test rate relative to total tests but it can be difficult to find explanations of how this metric is calculated. There was no national guidance on how to present such data (Dyrda and Drees, 2020). However, many states do report positive test rate relative to individuals tested, which is a better reflection of community case load than positive test rate relative to total tests conducted. States developed their COVID dashboards independently without any federal guidance, so they are all different in terms of their usability, detail of data, frequency of updates, ability for data to be downloaded, and more (Tracking COVID-19 in the United States: From Information Catastrophe to Empowered Communities, 2020).

### 2.1.3. Timing

Data reporting also faces challenges of timing. All daily datasets of COVID-19 information are cyclical: the numbers routinely dip lower every sixth and seventh day of reporting, representing less testing and fewer medical visits on weekends. Illness does not follow the seven-day week, but its record is defined by it. The challenge of the seven-day week also extends to questions of funerary processing; for example, a daily

cremation rate has little use in a model because crematories often do not run on the weekends. Processing metrics for human remains thus must be quantified on a weekly basis, not daily.

Although many COVID-19 data dashboards note the dates and times that data was collected and the dashboard updated, daily datasets may still contain an undisclosed lag between the true collection date and the date that the data is reported. For example, early in the pandemic, our research team discovered that one state website had a 3-day lag in hospital data (i.e., number of hospitalized patients with COVID-19 in general care and ICU). This meant that the COVID-19 patients reported on May 12, for example, actually represented the number of COVID-19 patients in hospitals on May 9. This data lag has improved, but contradictions between two data sources maintained by the state suggests that a two-day data lag still exists. Data lags are likely caused by the need to collect all relevant data for a particular day, check the data to make sure they are free of error, and transition the data to where it will be available to the general public, which is no small task. However, lack of clarity about the day that the data actually represents can hinder efforts at comparing models to real-world conditions to check the accuracy of the model which will be used to make decisions on resource allocation. Having data tagged with the date it actually represents is especially important during the early days of a new wave, when virus spread approaches an exponential rate and a couple of days of difference could underestimate by thousands of cases. Data that is reported on a weekly basis, such as COVID-19 cases in long term care facilities, offers less granularity but is likely more robust to lags.

Other timing issues influence data comparison through metrics such as the case fatality rate (CFR), meaning the fraction of people who contract COVID-19 that ultimately die. Johns Hopkins University’s COVID dashboard website, for example, presents the CFR as the ratio of cumulative deaths by the cumulative positive cases recorded on the same day (COVID, 19). Various researchers (Donnelly et al., 2003; Linton et al., 2020; Wilson et al., 2020; Yang et al., 2020) have noted that this can underestimate the true CFR, particularly when the pandemic is ongoing, since the cumulative case count would include people whose outcomes are yet to be determined. If the pandemic were over, this may not be an issue. But since the pandemic is ongoing, and the counts are constantly changing, this lag time must be accounted for. Additionally, our own explorations into CFR show a steady decrease over time, potentially owing to a combination of unknown variables such as improvements in treatment, reduced susceptibility of the remaining population and changing viral characteristics. If the CFR is indeed declining, cumulative death and case numbers that include high CFR periods from earlier in the pandemic are not necessarily helpful in calculating the current CFR.

Despite deaths being reported daily, and the media using terms like “largest one-day spike in deaths”, daily reported COVID-19 deaths rarely include deaths that have occurred that day. A coroner’s reported cause of death may come weeks after the actual death (Lu, 2020). This is because it takes time and staffing for a death to be certified, cross-referenced, and then finally tallied as a COVID-19 death. According to the director of the Virginia Department of Health, the death of a nursing-home resident with a known COVID-19 infection can be counted and reported almost immediately since they are already monitored by a health department worker, but other deaths can take up to 30 days to be reported as a COVID-19 death and added to a daily total (Harris, 2020). In contrast, in Florida, it takes closer to a week for deaths to be reported, demonstrating the potential for differences in this metric between states (Matthews, 2020).

Finally, states may update/improve their data by revising prior numbers as more information is received. However, aggregation sites that draw data from state websites may not be implementing similar revisions. Thus, the aggregated data may be less trustworthy over time than the sum of its parts.

## 2.2. Data gaps

In addition to imperfections in available data, there are data gaps about critical characteristics of the disease such as how long immunity lasts after an infected person has recovered, and whether vaccinated people can still transmit disease. This data does not yet exist because there has been limited time to study it. However, preliminary data is quickly emerging.

Other gaps include quantified metrics related to certain essential industries that have simply never been assembled for planning purposes, and are too disaggregated to meaningfully estimate based on a small sample size. Our research team encountered several of these critical data gaps, described below, while assembling information for New England's COVID-19 response. In the future, a central authority could coordinate a data gathering process among the disparate institutions to assure this information is available.

As infections mounted, policy makers strove to understand funerary systems, which ensure the dignified disposition of human remains, an important function for broader society. To estimate the capacity of the funerary system to process remains, our team combined cremation machine capacity per day and the number of cremation machines in each state to generate weekly cremation capacity. For New England, our team first collected data on crematories from the state registries of licensed crematory operators, but this data was only available for four of the six New England states. Furthermore, consultation with crematory operators revealed that this was still insufficient to estimate capacity, as the number of crematory machines in a single crematory facility varies. However, we subsequently discovered that organizations in each state had previously compiled the data on cremation machines by surveying individual crematories. Importantly, the organizations differed by state: in Connecticut, it was the State Medical Examiner's office; in Maine an individual crematorium manager had collected data for the whole state on his own initiative; in Massachusetts the data was collected by the private consulting group Cemetery Helpful Solutions; and in New Hampshire by the Public Health Preparedness Planner for the Department of Health and Human Services. Finally, in Rhode Island the Executive Director of the Rhode Island Funeral Directors Association held the data, and in Vermont it was the Vermont Agency of Natural Resources because of air quality concerns. Identifying these organizations was time consuming, but the compiled information existed. The same cannot be said for burial capacity, however, which derives from a) current available space in cemeteries and b) the equipment, which may be privately owned or owned by the municipality, needed to dig and fill in the graves, the availability of which will depend on the number of cemeteries served, the timing of the funerals, and available operators. Together this information proved too disaggregated to make any reliable predictions about a state's maximum capacity to bury human remains during a COVID-19 surge.

Our team also tried to estimate refrigeration capacity for human remains by state after the New York City funerary industry experienced delays in April. New York City was mostly able to compensate by providing hospital morgues with additional refrigeration capacity, while individual funeral homes also invested in auxiliary refrigeration. To determine whether New England states should invest in refrigerated trucks for hospitals, we would need to sum refrigeration capacity of hospital morgues, funeral homes, crematories, and the Medical Examiner's Office, while accounting for: 1) embalmed bodies do not require refrigeration, 2) not all human remains have an opportunity to be stored at hospitals (e.g. remains for nursing homes often go directly to funeral directors), 3) very few remains have an opportunity to be refrigerated at the Medical Examiner's Office, and 4) refrigeration demands will grow if there are funerary backlogs, which cannot be reliably estimated without maximum burial rates. Though we proposed an estimate for refrigeration time demand per human remain, refrigeration capacity proved too disaggregated to meaningfully estimate for the New England states.

Lastly, PPE burn rates refer to the amount of PPE used per worker per

unit time in order to protect themselves from contracting COVID-19 while they perform their work. Much policy attention focused on PPE supplies for hospital staff after reports of shortages early in the pandemic. By mid-August 2020, over 900 U.S. health care workers, including doctors, nurses, paramedics, and critical support staff such as nursing home workers and hospital custodians, had died from COVID-19 and its complications (JEMS, 2020), deaths that PPE should have protected them from. Yet the process that individual hospitals or other institutions use to calculate PPE burn rates is often unclear. Our team's work to advise on PPE allocations according to states' COVID-19 caseloads began by adapting an Ebola PPE model that calculated burn rates per medical staff. We quickly switched to using burn rate per patient per day because of data availability. Burn rates are different for nurses and doctors, as well as other staff that might come into contact with patients, but those roles and their specific rates of patient contact are not quantitatively defined and may significantly vary between different hospitals according to culture.

Furthermore, medical experts have observed that, for example, PPE utilization in the Boston South Shore Health System in March did not match utilization in April even when the numbers of staff and patients did not vary (Dyrda and Drees, 2020). Massachusetts had PPE shortages starting in March through at least June (Martin, 2020), and PPE burn rates change according to scarcity conditions – meaning that burn rates will be lower when materials are scarce, as workers are more likely to use them sparingly in order to avoid shortages, and also as standards continue to evolve (Dyrda and Drees, 2020). Changes in inventory can be estimated with reporting about PPE stocks, but this data is difficult to come by, because inventory tracking systems had not previously been designed to track this information even within hospital management, and because scarcity may lead to atypical sourcing of materials (Dyrda and Drees, 2020) including PPE that necessitate faster burn rates due to lower quality protection. Additional difficulties in incentivizing PPE reporting are explored in the following section.

These examples of data gaps are not new, but in the past they never obstructed providing necessary time-sensitive services. The system had enough excess capacity that material shortages could be avoided without anyone knowing precisely the throughput capacities. However, at a time when systems are pressured to provide service for an ever-growing number of patients and deceased remains, it becomes difficult to predict where the shortfalls or backlogs will emerge and to allocate resources accordingly.

In some cases of material shortages, independent actors may coordinate, like the many individual funeral home directors that coordinated across state lines to ensure timely cremations of the deceased. In other cases, like hospitals, some may manage to keep adequate PPE supplies while nearby hospitals substitute trash bags for hospital gowns, to the peril of their staff (Al-Arshani, 2020). Such disparities in material provisions do not serve society at large nor public health. Additionally, we note that medical providers are not the only critical demand for PPE: it is also needed for public shelters (e.g., homeless or natural disaster evacuation), public transit works, funeral home employees, grocery store workers, first responders, prisons, long-term care facilities, schools, and more. Estimates of PPE usage and requirements are increasingly difficult to determine as we move away from hospitals and towards these other facilities. These public services should be supported by adequate PPE, but without reliable information about usage in these facilities, it is difficult to calculate burn rates for planning and allocation decisions.

The challenges of data reporting thus take many forms and pose limits to characterizing the COVID-19 situation in the United States and strategizing an appropriate response. Again, emergency responses can be quicker and more efficient if data on key industries and populations are gathered into centralized, integrated, and accessible databases during times of non-emergencies (Luciano, 2020) and are updated with relevant new data as the emergency progresses (Sipior, 2020). Data aggregation, interpretation, and communication issues are considered in the next section.

### 3. Data aggregation and interpretation

#### 3.1. Data aggregation

There can be uncertainties embedded in the ways that data is aggregated. Because COVID-19 is a new disease, scientists and public health professionals are still trying to characterize it to predict population and resource trajectories in the future. This is not always straightforward due to outlier cases. For example, Cara Babachicos, Senior Vice President and CIO of South Shore Health System (Boston, MA) refers to the “fallacy of averages” when calculating the average length of COVID-19 patient stay in hospitals. Because a few outlier patients stayed a very long time, this skewed the average length of stay that planners used for everyone, creating a flawed and unhelpful statistic (Dyrda and Drees, 2020). The average length of stay of COVID-19 patients is used to calculate demand and availability of a number of hospital resources, including beds, ventilators, PPE, and medical staff. Having a realistic and reliable estimate is imperative to ensure healthcare facilities are equipped to care for their patients, but has been difficult to discern owing to the novelty of the virus. When calculating disease characteristics, care should be taken to identify and resolve outliers. This could mean using a median instead of an average, removing outliers from the dataset, or reporting the standard error along with the average. A single unified source of guidance for reporting common COVID-19 characteristics and their calculations would help ensure that values can be reliably used by hospitals and compared between hospitals, states, or regions.

Additionally, states report a positive test rate for COVID-19 as the ratio of the number of positive tests to the total number of tests conducted. The fraction itself is thus strongly influenced by the prevalence of testing. For example, some states may only have capacity to test symptomatic people, while others host colleges that have privately-funded robust testing programs that greatly enlarge the denominator. On October 12, Mississippi reported a test positivity rate of 100 %, but the state tested only 2 out of every 10,000 people. In contrast, Massachusetts had the second lowest test positivity rate in the nation of 0.9 %, but had tested 92 out of every 10,000 people, more than double the testing rate of any other state except North Dakota (which tested 91 out of every 10,000 people) and 46 times Mississippi’s testing rate (Gamble, 2020). Massachusetts Governor Charlie Baker acknowledged in mid-October that as the rate of testing has greatly increased since early in the pandemic, a positive test rate from total tests conducted has become less meaningful and suggested the state Department of Public Health would revise its dashboard accordingly (Anderson, 2020). Thus, the test positivity rate alone, without information about the denominator or overall population, has limited use in comparing states or even comparing one state over time. More informative data might include the identified cases per capita, and the overall rate of testing.

#### 3.2. Mandates, interpretability, and access rights of data reporting

Data does not gather itself. First a person or institution must pose a question and then implement a methodology that samples the subject of inquiry in a meaningful way, whether through surveys, automated algorithms that can comb and parse online information, manual observations, or other methods. These efforts produce a dataset that is maintained, and effectively owned, by the person or institution that gathered it. Data gathering efforts can be considerable, requiring time, effort, money, usage of information technologies, and other resources. Establishing data systems during, rather than before, an incident adds burden to organizations simultaneously carrying out critical emergency

response services.

Accessibility and costs of obtaining data after it is collected can vary. For example, annual cremation statistics are compiled into a detailed report by the Cremation Association of North America, with the full report only accessible to paying members (CANA, 2020).<sup>1</sup> The reasons for such restrictions are usually not stated, but could include recouping the costs of gathering it, justifying membership in trade organizations, or a concern that publishing the data would undermine the needs of the organization. The last reason may have been the case for hospitals early in the pandemic when private hospitals and other healthcare facilities displayed reluctance to publish or share information on PPE inventory, usage rates, and other data needed for efficient resource allocation. Sharing that they had even a small reserve might deprioritize their hospital to receive auxiliary supplies that they knew they would need in the near future. It has been suggested that the competitive environment of the healthcare industry caused this data to be treated as proprietary information (Evans and Berzon, 2020), creating a significant obstacle for data gathering. Eventually, hospitals and long-term care facilities were required to report days-on-hand of critical PPE supplies to the U.S. Department of Health and Human Services (HHS) (King, 2020). There were low rates of reporting in some states during the first few weeks after HHS reporting requirements took effect, making it difficult to properly assess capacity, but compliance has since risen so that greater than 87 % of hospitals in each state report data to HHS on a weekly basis (HHS Protect Public Data Hub, 2020).

The limits of data sharing manifest in other ways as well, including through government mandates. In Massachusetts, the Department of Public Health (DPH) gathers data on positive COVID-19 cases, but claimed a privacy exemption from the public records law to avoid providing data about daycare programs with positive tests, despite making public statements summarizing the data (Ebbert, 2020a). It was not until September 2020, following two appeals from the Boston Globe, that the state released case information for individual child-care centers, making it possible to track potential transmission among children (Ebbert, 2020b). At the time of writing, DPH is providing raw weekly data on positive COVID-19 cases in family daycares/small group programs by municipality and county on its website under the title Chapter 93 EEC Weekly Report (Massachusetts Department of Public Health, 2020). The title, which mentions neither “child” nor “care”, makes the data difficult to find through searches, and the dataset itself does not include the individual center information that was provided to the Boston Globe, which would make it possible to draw conclusions about outbreaks. Similarly, Massachusetts state senators had to push for data regarding COVID-19 outbreaks in long-term care facilities even though the governor had signed a law requiring DPH to provide this data in June 2020 (Schoenberg, 2020). The governor claimed that data reporting for long-term care facilities is sometimes insufficient for the disaggregation requested (Schoenberg, 2020), but this was not the case for daycare data, since the state provided generalized information about center-specific outbreaks in other instances. Other reasons for failing to provide public data can include insufficient staffing to process the data, although staffing assignments invariably reflect the priorities of their managers.

Individual state record keepers can also make mistakes in data presentation. For example, in autumn 2020, the Massachusetts COVID-19 dashboard displayed a y-axis for “new hospitalizations count” when they were actually plotting the daily change in the number of patients hospitalized. That graph is now accurately labeled as “changes in confirmed hospitalized patients by date.”

Public health data may also be provided through non-governmental sources, raising questions about its provenance and veracity. In April

<sup>1</sup> We note that during our research on cremation capacity, the Executive Director of the Cremation Association of North America was happy to share the data for specific questions to representatives of U.S. Army Corps of Engineers.

2020 and potentially later, the U.S. Centers for Disease Control and Prevention (CDC) reported confirmed cumulative COVID-19 case count data from USAFacts (Lasry et al., 2020; USAFacts, 2020), a privately-funded organization that presents itself as a non-partisan, not-for-profit civic initiative without a political agenda. It “rel[ies] solely on government data for consistency and to screen for bias” but does not provide information on data processing, though the website encourages questions and communications (USAFacts, 2020). While there is no reason to suspect data manipulation, the accountability of such a private entity is markedly different from public government data. In cases where non-governmental sources of data are used for official government reporting, justification should be provided.

USAFacts proved particularly useful in part because the US government stalled to implement a consistent mechanism for collecting COVID-19 data. Several months into the emergency, HHS eliminated the option for hospitals to submit their required COVID data to the CDC, forcing them to use recently developed tools. At the same time, the number of required data elements greatly increased and the types of hospitals required to report were expanded to include, for example, rehabilitation, oncology, orthopedic, and psychiatric (Goldstein and Sun, 2020). Hospitals were expected to comply with the new requirements only a few days after they were announced. This disrupted the reporting system that had been working smoothly in some states, creating a backlog in data, making it more difficult for states to receive data on their own hospitals, and potentially interfering with patient care as the hospitals struggled with the transition (McKenna, 2020). Hospitals that were less digitally mature (i.e. lacked robust digital information management systems and personnel) were likely more impacted by changing requirements as they lacked the flexibility of more digitally mature hospitals (Dwivedi et al., 2020; Fletcher and Griffiths, 2020). Miscommunications continued into August, with conflicting information about whether the CDC would again gain control of the data, and the CDC’s involvement in the decision-making process (Jercich, 2020). The American Hospital Association advised its members to comply with the requirements to ensure they were appropriately prioritized for medicine distribution (Jercich, 2020), a policy that effectively compels their participation in an imperfect system. A new emergency rule in August from the Centers for Medicare and Medicaid Services additionally enforced compliance by terminating Medicare and Medicaid funding for hospitals that do not follow the reporting requirements (Stolberg, 2020).

The lack of a well-established, consistent pipeline and clear guidance for reporting have created discrepancies between sources of data. Shortly after the reporting requirements changed, the COVID Tracking Project, a “volunteer organization launched by *The Atlantic* and dedicated to collecting and publishing [COVID-19] data” compared hospitalization data reported by the federal government and state health departments. They found that HHS was reporting 24 % more hospitalized patients nationwide than states were the week ending July 26 (Glassman and Ladyzhets, 2020). This difference between federal and state hospitalizations varied by state, with HHS counts within 5 % of Arizona counts but well exceeding 200 % of the count reported by Wisconsin. The HHS data also reported large and sudden spikes or dips in hospitalizations that were not present in the state data. While these discrepancies may be partially explained by which patients get counted (HHS counts confirmed and suspected cases of COVID-19 in hospitals while some states only count confirmed cases), the COVID Tracking Project found that differences in definition alone did not explain the variability.

However, some of these data collection concerns have been alleviated. In February 2021, the COVID-19 Tracking Project announced they would close operations by March 2021 after observing “persuasive evidence that the CDC and HHS are now both able and willing to take on the country’s massive deficits in public health data infrastructure, and to offer the best available data and science communication in the interim” (Kissane and Madrigal, 2021). Furthermore, the announcement to close

operations states that the work of compiling, cleaning, standardizing, and making sense of COVID-19 data is properly the work of federal public health agencies, both because these efforts are a governmental responsibility and because federal teams have access to far more comprehensive data and can mandate compliance with at least some standards and requirements (Kissane and Madrigal, 2021).

### 3.3. Access to and consistency of interpretations

Like the data itself, data interpretations may not be freely accessible to the public in all circumstances. Newspapers, while shifting from traditional print form to online form, have implemented paywalls for general content. A number of newspapers and journalistic websites suspended their paywall for articles concerning COVID-19. This can present a long-term problem for funding the type of investigative journalism that provides in-depth, vetted, and timely information, and some organizations later reinstated their blanket paywall to include COVID-19 coverage (Saltz, 2021). However, the alternative, a world in which well-researched data interpretation articles cost money while many media sources or individuals without the investigative capabilities provide lower-standard coverage for free, is ripe to drive attention towards information that is less accurate and less helpful in protecting public health. For example, in a study on Americans’ ability to identify fake information related to COVID-19, researchers found that 20–25 % of respondents believed fake information but that belief decreased to varying extents when corrections on the fake information were published (Kreps and Kriner, 2020). Alarming reports and messages generally gain more attention, but it is important to have a balance of alarming and reassuring content from both the government and news media to foster an engaged but peaceful society (Rao et al., 2020). Thus, the availability and tone of high-quality data interpretation can have direct impacts on behaviors and the pandemic’s subsequent infection trajectory.

Additionally, there is the issue of inconsistent interpretations – both inconsistent in their reflection of the data, and inconsistent over time. The CDC has had several reversals of recommendations. First, in February 2020, the CDC published that it did not recommend wearing facemasks to prevent COVID-19 transmission (Buchwald, 2020). This guidance was not motivated by a lack of scientific evidence for the effectiveness of masks in general, but rather by a concern that the general public did not know how to wear the masks properly, making them ineffective and thus wasting scarce resources that were desperately needed by healthcare workers (Weintraub, 2020). The recommendation was reversed in early April (Miller and Stobbe, 2020) and has since been further reinforced by scientific studies (CDC Media Relations, 2020). This now-famous reversal, which was also observed in other countries (Janssen and van der Voort, 2020), is an understandable part of the evolving process of managing a public health emergency as organizations strive to ensure that materials, such as medical-grade facemasks, go to where they are needed most as well as make recommendations based on scientific findings. Other CDC reversals have included whether COVID-19 can be transmitted through air (Elfrink et al., 2020) and whether asymptomatic contacts of positive cases should be tested (Hellmann, 2020). The explanations for these reversals were attributed, respectively, to website error and political interference (Stieb, 2020). The idea that public health recommendations are a function of political pressure or technical mistakes rather than the best available science will erode trust by the very people who most need the information. It is one thing to incorporate new information and learn, but another to issue recommendations that are not based on the best available information.

## 4. Implications for research

This document has described the pitfalls encountered in our team’s research in New England, but has not provided a comprehensive assessment of data pitfalls related to coronavirus data management, nor

data management in other fields. There is extensive opportunity to identify inefficiencies, gaps, and ambiguities in data management and data compilation across fields, as well as to learn lessons from successful data management practices. The coronavirus pandemic has demonstrated the necessity for accessing and applying data quickly, and can be supported through researcher efforts to advocate for good data practices outside of emergencies to ensure they are established when the emergencies occur.

## 5. Recommendations

Herein we provide individual actions that can make data gathering and interpretation more efficient, and broad-scale institutional actions, ideally led by federal initiatives. These recommendations are parsed into those useful to researchers, who might include data managers, data collectors, and analysts, and those that can only be applied through the institutions managing data collections or the data gathering structures.

### 5.1. Researcher best practices

A lesson we learned while gathering data about crematory machines is that a comprehensive survey can be avoided if you make a few calls while asking the question “has anyone else called you to collect this data?” This helped us identify which organizations had already aggregated all the information and prevented us from having to reinvent the wheel.

When designing and assembling data spreadsheets or numerical reports for public use, define the terms and their mathematical derivations, ideally after researching how other similar reports are using and defining terms. These reports should be cited when describing your own definitions. Be cognizant that other fields may use language differently, so clarity is of utmost importance. Data dictionaries are highly appreciated.

When collecting or aggregating data, attach detailed notations of sources, definitions, dates, and other relevant explanations; when utilizing data, look for such notations.

Data closest to the source is likely to be most reliable. The people or institutions that have directly collected the data have the best knowledge of what was included (e.g., which types of ventilators were counted in their ventilator supply), what different terms refer to (e.g., whether counts of gloves refer to single, pairs, or boxes of gloves), and how different variables were calculated (e.g., burn rates or length of hospital stay). The more the data is passed through various organizations and the more it is aggregated, the greater the chances there are for errors to be introduced and original notes to be lost.

### 5.2. Institutional best practices

The foremost need during an unfolding emergency is for a centrally-recognized authority to issue standards for all data reporting entities to follow. This would preclude the need for each individual entity to research and define terms, as well as for data users to review the usage of the terms. A standard set of linguistic and mathematic definitions will save everyone time. Data can then be easily aggregated without conflating different measurements in the same value.

Data reporting should be similarly streamlined. The date to which data applies (e.g., frequency and last date and time of data collection) should be documented consistently and in a manner that is easily visible to the reader.

Encourage data collection for critical services, like fatality management, PPE provisions, and hospital bed capacity, during calm times so there is a level for comparison during shortages.

Provide free access to data and interpretations. Many states, the CDC, and other organizations post raw data and reports that are easily downloadable from their websites.

Be responsive to user needs. State and federal agencies and other

organizations have periodically revised their data dashboards and published additional reports as the COVID-19 pandemic has progressed and they received feedback from users.

Where data is imperfect or contains inconsistencies, clearly indicate this in the reporting. For example, the New York Times profiles of states' COVID-19 data includes a section at the bottom entitled “About the Data” that lists reporting anomalies or methodology changes in the data (see [The New York Times, 2020](#)).

Various state websites contain subsets of data on topics such as the types of settings in which outbreaks have occurred, and cases in public schools, colleges, and universities, which are informative for policy decisions, epidemiological research, and community members' personal decisions. Providing these disaggregations of data, where possible, can deepen the breadth of analysis.

Recommendations provided to the public should derive from sound science and ideally be consistent across organizations, both of which should be easy to verify. If public trust—an invaluable commodity to leaders during a pandemic—is to be preserved, guidance regarding safety and healthcare should be clearly distinguished from guidance motivated by resource management concerns. The role that each institution plays in responding to the pandemic should be clearly delineated in this context to increase transparency and preserve public trust. Scientific facts are indifferent to our resource challenges. The guidance of scientific institutions entrusted to inform the public for their safety and health should always communicate based on that mandate.

## 6. Conclusions

This paper has classified and reviewed different types of data challenges that emerged or gained more attention during the progression of the coronavirus pandemic, beginning in March 2020. The field of epidemiology, and public health in general, was cognizant of some of its limitations prior to the pandemic, but the consequences of inadequate information and action are greater during an unfolding national and international emergency. This has presented data managers and institutions alike with reason to reflect upon and potentially reorder structures and practices that could save time in managing future events. Data management can be improved, and the lessons learned from the COVID-19 pandemic should galvanize implementing improvements in the near future. This document has presented several options for action.

### Author statement

**Stephanie Galaitsi:** Writing - Original draft preparation, analysis. **Jeff Cegan:** Conceptualization, Writing- Reviewing and Editing, **Kaitlin Volk:** Writing- Reviewing and Editing, **Matthew Joyner:** Writing- Reviewing and Editing, **Benjamin Trump:** Methodology, **Igor Linkov:** Conceptualization.

### Disclaimer

The views expressed in the article do not necessarily represent the views of the U.S. Army Corps.

### Role of funding source

This paper was funded by general funds, and not out of any particular project.

### Declaration of Competing Interest

The authors declare no competing interests.



## References

- Al-Arshani, S. (2020). Nurse dies in New York hospital where workers are reduced to using trash bags as protective medical gear. March 26. Business Insider <https://www.businessinsider.com/kious-kelly-hospital-nurse-dies-trash-bags-2020-3>.
- Anderson, K. (2020). Massachusetts to make changes to COVID-19 reporting. WCVB5 (October 20) 11/06/2020.
- Barnes, S. J. (2020). Information management research and practice in the post-COVID-19 world. *International Journal of Information Management*, 55, Article 102175.
- Buchwald, E. (2020). U.S. health officials say Americans shouldn't wear face masks to prevent coronavirus—Here are 3 other reasons not to wear them. March 2. Market Watch <https://www.marketwatch.com/story/the-cdc-says-americans-dont-have-to-wear-face-masks-because-of-coronavirus-2020-01-30>.
- Bunker, D. (2020). Who do you trust? The digital destruction of shared situational awareness and the COVID-19 infodemic. *International Journal of Information Management*, 55, Article 102201.
- CANA. (2020). Industry statistical information. All things cremation. <https://www.cremationassociation.org/page/IndustryStatistics>.
- CDC. (2020a). COVID-19 pandemic planning scenarios. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html>.
- CDC. (2020b). Weekly updates by select demographic and geographic characteristics: Provisional death counts for coronavirus disease 2019 (COVID-19). Centers for Disease Control. [https://www.cdc.gov/nchs/nvss/vsrr/covid\\_weekly/index.htm?fbclid=IwAR3-wrg3tTKK5-9tOHPGAHWFO3DfslkJK0KsDEPQpWmPbKtp6EsoVV2QsIQ](https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm?fbclid=IwAR3-wrg3tTKK5-9tOHPGAHWFO3DfslkJK0KsDEPQpWmPbKtp6EsoVV2QsIQ).
- CDC Media Relations. (2020). CDC calls on Americans to wear masks to prevent COVID-19 spread. CDC. <https://www.cdc.gov/media/releases/2020/p0714-americans-to-wear-masks.html>.
- Colaneri, M., Seminari, E., & Novati, S. (2020). COVID19 IRCCS San Matteo Pavia Task Force. Severe acute respiratory syndrome coronavirus 2 RNA contamination of inanimate surfaces and virus viability in a health care emergency unit. *Clinical Microbiology and Infection*, 26(1094), e1–e1094.
- Colaneri, M., Seminari, E., Piralla, A., Zuccaro, V., Di Filippo, A., Baldanti, F., Bruno, R., Mondelli, M. U., Brunetti, E., & Di Matteo, A. (2020). Lack of SARS-CoV-2 RNA environmental contamination in a tertiary referral hospital for infectious diseases in Northern Italy. *The Journal of Hospital Infection*.
- COVID, C. (2020). Dashboard by the center for systems science and engineering (CSSE) at Johns Hopkins University (JHU) (19).
- Donnelly, C. A., Ghani, A. C., Leung, G. M., Hedley, A. J., Fraser, C., Riley, S., Abu-Raddad, L. J., Ho, L.-M., Thach, T.-Q., & Chau, P. (2003). Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *The Lancet*, 361(9371), 1761–1766.
- Doyle, R., & Conboy, K. (2020). The role of IS in the covid-19 pandemic: A liquid-modern perspective. *International Journal of Information Management*, 55, Article 102184.
- Dwivedi, Y. K., Hughes, D. L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J. S., Gupta, B., Lal, B., Misra, S., & Prashant, P. (2020). Impact of COVID-19 pandemic on information management research and practice: Transforming education, work and life. *International Journal of Information Management*, 55, Article 102211.
- Dyrda, L., & Drees, J. (2020). The COVID-19 data issues 12 health system IT execs want addressed first. Becker's Health IT. <https://www.beckershospitalreview.com/health-care-information-technology/the-covid-19-data-issues-12-health-system-it-exec-wants-addressed-first.html>.
- Ebbert, S. (2020a). State won't divulge COVID-19 cases in child care centers. August 10. The Boston Globe <https://www.bostonglobe.com/2020/08/10/metro/state-wont-divulge-covid-19-cases-child-care-centers/>.
- Ebbert, S. (2020b). State's emergency child-care centers kept COVID-19 in check. Here's how. September 22. The Boston Globe <https://www.bostonglobe.com/2020/09/22/metro/states-emergency-child-care-centers-kept-covid-19-check-heres-how/>.
- Elfrink, T., Guarino, B., & Mooney, C. (2020). CDC reverses itself and says guidelines it posted on coronavirus airborne transmission were wrong. September 21. Washington Post <https://www.washingtonpost.com/nation/2020/09/21/cdc-covid-aerosols-airborne-guidelines/>.
- Evans, M., & Berzon, A. (2020). Why hospitals can't handle Covid surges: They're flying blind. September 30 <https://www.wsj.com/articles/hospitals-covid-surge-data-11601478409>.
- Fahey, R. A., & Hino, A. (2020). COVID-19, digital privacy, and the social limits on data-focused public health responses. *International Journal of Information Management*, 55, Article 102181.
- Fairchild, G., Tasseff, B., Khalsa, H., Generous, N., Daughton, A. R., Velappan, N., Priedhorsky, R., & Deshpande, A. (2018). Epidemiological data challenges: Planning for a more robust future through data standards. *Frontiers in Public Health*, 6, 336.
- Fletcher, G., & Griffiths, M. (2020). Digital transformation during a lockdown. *International Journal of Information Management*, 55, Article 102185.
- Gamble, M. (2020). States ranked by COVID-19 test positivity rates: Oct. 13. Becker's Hospital Review. <https://www.beckershospitalreview.com/public-health/states-ranked-by-covid-19-test-positivity-rates-july-14.html>.
- Glassman, R., & Ladyzhets, B. (2020). Hospitalization data reported by the HHS vs. the states: Jumps, drops, and other unexplained phenomena. August 11. The COVID Tracking Project <https://covidtracking.com/blog/hospitalization-data-reported-by-the-hhs-vs-the-states-jumps-drops-and-other>.
- Goldberg, C. (2020). Big mass. Study finds you could be symptom-free but still carry high load of coronavirus. August 6. WBUR <https://www.wbur.org/commonhealth/2020/08/06/broad-nursing-homes-coronavirus-viral-loads>.
- Goldstein, A., & Sun, L. H. (2020). Hospital officials, experts say new federal rules for covid-19 reporting will add burdens during pandemic. July 15. The Washington Post <https://www.washingtonpost.com/health/2020/07/15/coronavirus-trump-administration-data-change/>.
- Harris, S. (2020). VDH explains backlog in COVID-19 death reporting. September 18. WAVY / WVBT <https://www.wavy.com/news/health/coronavirus/vdh-explains-backlog-in-covid-19-death-reporting/>.
- Hellmann, J. (2020). CDC reverses guidance on COVID-19 testing for asymptomatic people. September 18. The Hill <https://thehill.com/policy/healthcare/517091-cdc-reverses-guidance-on-covid-19-testing-for-asymptomatic-people>.
- HHS Protect Public Data Hub. (2020). U.S. hospital reporting dashboard. U.S. Department of Health and Human Service. <https://protect-public.hhs.gov/pages/covid19-mod-ule>.
- Janssen, M., & van der Voort, H. (2020). Agile and adaptive governance in crisis response: Lessons from the COVID-19 pandemic. *International Journal of Information Management*, 55, Article 102180.
- JEMS. (2020). More than 900 U.S. medical staff have died fighting COVID-19. *Journal of Emergency Medical Services*. <https://www.jems.com/2020/08/12/more-than-900-us-medical-staff-died-covid-19/>.
- Jercich, K. (2020). HHS said to reverse course, with hospitals to report COVID-19 data to CDC again. August 20. Healthcare IT News <https://www.healthcareitnews.com/news/hhs-said-reverse-course-hospitals-report-covid-19-data-cdc-again>.
- King, R. (2020). CMS to require hospitals to report critical COVID-19 data on bed capacity, PPE and cases. August 25. Fierce Healthcare <https://www.fiercehealthcare.com/hospitals/cms-to-require-hospitals-to-report-critical-covid-19-data-bed-capacity-ppe-and-cases>.
- Kissane, E., & Madrigal, A. (2021). It's time: The COVID tracking project will soon come to an end. February 1. The Atlantic <https://covidtracking.com/analysis-updates/covid-tracking-project-ended-march-7>.
- Kreps, S. E., & Kriner, D. (2020). Medical misinformation in the COVID-19 pandemic. Available at SSRN 3624510.
- Lasty, A., Kidder, D., Hast, M., Poovey, J., Sunshine, G., Winglee, K., Zviedrite, N., Ahmed, F., & Ethier, K. A. (2020). Timing of community mitigation and changes in reported COVID-19 and community mobility – four U.S. metropolitan areas, February 26–April 1, 2020. Centers for Disease Control. <https://www.cdc.gov/mmmwr/volumes/69/wr/mm6915e2.htm>.
- Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S., Yuan, B., Kinoshita, R., & Nishiura, H. (2020). Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *Journal of Clinical Medicine*, 9(2), 538.
- Lu, D. (2020). The true coronavirus toll in the U.S. has already surpassed 200,000. August 12. New York Times <https://www.nytimes.com/interactive/2020/08/12/us/covid-deaths-us.html>.
- Luciano, E. M. (2020). Information management hits and misses in the COVID19 emergency in Brazil. *International Journal of Information Management*, 55, Article 102194.
- Madrigal, A. C., & Meyer, R. (2020). How could the CDC make that mistake?. May 21. The Atlantic <https://www.theatlantic.com/health/archive/2020/05/cdc-and-states-are-misreporting-covid-19-test-data-pennsylvania-georgia-texas/611935/>.
- Mandavilli, A. (2020). Your coronavirus test is positive. Maybe it shouldn't be. New York Times (September 17) <https://www.nytimes.com/2020/08/29/health/coronavirus-testing.html#:~:text=One%20solution%20would%20be%20to,37%2C%20to%20detect%20the%20virus>.
- Martin, N. (2020). It's like pulling teeth: There's still a PPE shortage—and a second wave could send medical workers into crisis mode. June 21. The Boston Globe <https://www.bostonglobe.com/2020/06/21/nation/dangerous-shortages-protective-gear-persist-mass-hospitals-clinicians-say/>.
- Massachusetts Department of Public Health. (2020). Archive of chapter 93 COVID-19 data. Massachusetts Department of Public Health. <https://www.mass.gov/info-details/archive-of-chapter-93-covid-19-data>.
- Matthews, J. (2020). COVID death toll looks "in the rear-view mirror," health experts say. July 10. Fox 13 <https://www.fox13news.com/news/covid-death-toll-looks-in-the-rear-view-mirror-health-experts-say>.
- McGrory, K., Woolington, R., & Humburg, C. (2020). Florida released a list of coronavirus deaths. But key information is blacked out. May 6. Tampa Bay Times <https://www.tampabay.com/news/health/2020/05/06/under-pressure-florida-released-a-controversial-list-of-coronavirus-deaths-but-key-information-is-blacked-out/>.
- McKenna, M. (2020). Covid-19 data in the US is an 'information catastrophe'. July 30. WIRED Magazine <https://www.wired.com/story/covid-19-data-in-the-us-is-an-information-catastrophe/>.
- Meyerowitz, E. A., Richterman, A., Gandhi, R. T., & Sax, P. E. (2020). Transmission of SARS-CoV-2: A review of viral, host, and environmental factors. *Annals of Internal Medicine*.
- Miceli, S. (2019). Exploring the complications of counting casualties after natural disasters. The National Academies of Sciences, Engineering, Medicine. <https://www.nationalacademies.org/news/2019/09/exploring-the-complications-of-counting-casualties-after-natural-disasters>.
- Miller, Z., & Stobbe, M. (2020). Face masks recommended, though Trump says he won't wear one. April 3. CBS2 KUTV <https://kutv.com/news/coronavirus/white-house-moves-toward-promoting-face-masks-to-fight-virus>.
- Mondelli, M. U., Colaneri, M., Seminari, E. M., Baldanti, F., & Bruno, R. (2020). Low risk of SARS-CoV-2 transmission by fomites in real-life conditions. *The Lancet Infectious Diseases*.
- Pearce, K. (2020). Excess death show the true impact of COVID-19 in the U.S. Johns Hopkins University. <https://hub.jhu.edu/2020/09/01/comorbidities-and-coronavirus-deaths-cdc/>.
- Rao, H. R., Vemprala, N., Akello, P., & Valecha, R. (2020). Retweets of officials' alarming vs reassuring messages during the COVID-19 pandemic: Implications for crisis management. *International Journal of Information Management*, 55, Article 102187.

- Riddell, S., Goldie, S., Hill, A., Eagles, D., & Drew, T. W. (2020). The effect of temperature on persistence of SARS-CoV-2 on common surfaces. *Virology Journal*, 17 (1), 1–7.
- Saltz, H. (2021). *Removing paywalls on coronavirus coverage is noble. It also makes no sense* (n.d.) Retrieved October 14, 2020, from. Poyter <https://www.poynter.org/business-work/2020/removing-paywalls-on-coronavirus-coverage-is-noble-it-also-makes-no-sense/>.
- Schoenberg, S. (2020). *COVID database filled with holes despite law*. July 10. Commonwealth Magazine <https://commonwealthmagazine.org/health-care/covid-database-filled-with-holes-despite-law-2/>.
- Sipior, J. C. (2020). Considerations for development and use of AI in response to COVID-19. *International Journal of Information Management*, 55, Article 102170.
- Stieb, M. (2020). *Pence says CDC will change school-reopening guidelines after trump complaint*. July 8. New York Magazine <https://nymag.com/intelligencer/2020/07/pence-cdc-change-school-guidelines-after-trump-complaint.html>.
- Stolberg, S. G. (2020). *Trump administration orders hospitals to report data or risk losing funding*. August 25. New York Times <https://www.nytimes.com/2020/08/25/us/politics/trump-coronavirus-data.html>.
- Taubes, G., & Mann, C. C. (1995). Epidemiology faces its limits. *Science*, 269(5221), 164.
- The New York Times. (2020). *Massachusetts Covid map and case count*. November 6. The New York Times <https://www.nytimes.com/interactive/2020/us/massachusetts-coronavirus-cases.html>.
- Tracking COVID-19 in the United States: From Information Catastrophe to Empowered Communities. (2020). *Prevent epidemics*. [https://preventepidemics.org/wp-content/uploads/2020/07/RTSL\\_Tracking-COVID-19-in-the-United-States\\_-7-23-2020.pdf](https://preventepidemics.org/wp-content/uploads/2020/07/RTSL_Tracking-COVID-19-in-the-United-States_-7-23-2020.pdf).
- USAFACTS. (2020). *US coronavirus cases and deaths*. USAFACTS. <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>.
- Walker, A. S., Jones, L. W., & Gamio, L. (2020). *Is the coronavirus death tally inflated? Here's why experts say no*. June 19. New York Times <https://www.nytimes.com/interactive/2020/06/19/us/us-coronavirus-covid-death-toll.html>.
- Wamsley, L. (2020). *Fired Florida data scientist launches a coronavirus dashboard of her own*. June 14. NPR <https://www.npr.org/2020/06/14/876584284/fired-florida-data-scientist-launches-a-coronavirus-dashboard-of-her-own>.
- Weintraub, K. (2020). *Not just ventilators: Staff trained to run them are in short supply*. *Scientific American*. <https://www.scientificamerican.com/article/not-just-ventilator-staff-trained-to-run-them-are-in-short-supply/>.
- Wilson, N., Kvalsvig, A., Barnard, L. T., & Baker, M. G. (2020). Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality. *Emerging Infectious Diseases*, 26(6), 1339.
- World Health Organization. (2009). *Pandemic influenza preparedness and response: A WHO guidance document*. World Health Organization.
- Yang, X., Yu, Y., Xu, J., Shu, H., Liu, H., Wu, Y., Zhang, L., Yu, Z., Fang, M., & Yu, T. (2020). Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: A single-centered, retrospective, observational study. *The Lancet Respiratory Medicine*.