Epidemic Tracking and Forecasting: Lessons Learned from a Tumultuous Year

Roni Rosenfeld[a] Ryan J. Tibshirani[a,b]

[a]Machine Learning Department, Carnegie Mellon University
[b]Department of Statistics & Data Science, Carnegie Mellon University

1 Introduction

Epidemic forecasting has garnered increasing interest in the last decade, nurtured and scaffolded by various forecasting challenges organized by groups within the U.S. federal government, including the CDC [1, 2, 3], OSTP [4], DARPA [5], and elsewhere [6, 7]. In 2017, after several years of experimentation with flu forecasting in academic groups, the CDC decided to incorporate influenza forecasting into its normal operations, including weekly public communications citepFluSight and briefing to higher-ups. To provide more reliable infrastructure and support for its forecasting needs, the CDC in 2019 designated two national Centers of Excellence for Influenza Forecasting, one at the University of Massachusetts at Amherst[1] and one at Carnegie Mellon University[2].

Not unrelatedly, the last decade has also seen a rise in the importance of *digital surveillance* streams in public health, with improving epidemic tracking and forecasting models being a key application of these data. Digital streams, such as search and social media trends, have constituted a large part of the focus [8, 9, 10, 11, 12, 13]; however, even more broadly, data from *auxiliary* streams that operate outside of traditional public health reporting, such as online surveys, medical devices, or electronic medical records, have received considerable attention as well [14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24].

The Carnegie Mellon Delphi group, which the two of us co-lead, has worked in both of these emerging disciplines—epidemic forecasting, and building relevant auxiliary signals to aid such forecasting models—since 2012. In 2020, as the pandemic broke out, we struggled like many other groups to find ways to contribute to the national efforts to respond to the pandemic. We ended up shifting our focus to lie nearly entirely on the data end of the spectrum, pursuing several different directions in order to build and make available to the public a variety of new indicators that reflect real-time COVID-19 activity in the U.S. Three papers in this collection describe this work from three different perspectives. A fourth describes international work by some of our collaborators that parallels our group's work on online surveys in the U.S.

2 Papers in This Collection

Here is a very brief summary of the papers in this collection.

1. Reinhart et al. [25] describes our group's (ongoing) effort in building and maintaining COVIDcast: an open repository of real-time, geographically detailed COVID-19 indicators in the U.S. These indicators (a term we use interchangeably with signals) are derived from a diverse set of data sources: medical testing devices, medical insurance claims, internet search trends, app-based mobility data, and online surveys, among others. Many indicators are demonstrated to have meaningful statistical relationships with what have become the pandemic's "topline" numbers (reported cases, hospitalizations, deaths), whereas others uniquely reflect certain activity (not available in other publicly-available data sources) that may drive or affect the spread of COVID-19. The paper demonstrates through a number of examples that the indicators in the COVIDcast repository can improve on the timeliness, robustness, and scope of traditional public health reporting streams.

[1]https://reichlab.io/people
[2]https://delphi.cmu.edu/about/center-of-excellence/

2. McDonald et al. [26] provides a detailed analysis of whether a core set of the indicators in the COVIDcast repository can be used to improve the accuracy of COVID-19 short-term forecasting and hotspot detection models. This speaks to the quantitative utility of the indicators in a way that is directly tied to the benefits observed in relevant downstream modeling tasks. The paper finds that time series models (that are already competitive with top forecasters from the COVID-19 Forecast Hub [27]) improve in predictive accuracy when they are supplemented with any of the five indicators under consideration. However, the magnitude of these improvements varies with the pandemic's local dynamics at prediction time, and only one indicator appears to offer a nontrivial improvement during upswings.

3. Salomon et al. [28] focuses on the U.S. COVID-19 Trends and Impact Survey (CTIS), an (ongoing) online survey operated by our group, in partnership with Facebook. This is a very rich source of data about the pandemic and its effect on people, only partially reflected by the indicators (derived from the survey) in the COVIDcast repository; the full data set of individual, anonymized survey responses is available to researchers under a data use agreement. The paper presents descriptive analyses that reflect the unique value of CTIS as an important supplement to public health reporting, in particular, as an instrument to measure key information about behaviors, attitudes, economic impacts, and other topics not covered in traditional public health streams.

4. *Authors' note to the Editor: this is to be filled out later. The Astley et al. paper is still undergoing a significant revision and its main messages may change as a result.*

3 Lessons Learned

We now take the opportunity to reflect on some "lessons learned" from our work over the past year and a half. Some of the observations below are described in more depth in the papers in the collection, and others extend beyond the papers in the collection (but we give references to relevant articles with more details in the discussion below).

Deceptively simple data labels often belie the data's true meaning and complexity. Labels such as "COVID-19 cases" or "COVID-19 hospitalizations" hide an enormous amount of complexity and potential ambiguity, especially when applied to data at fine geographic and temporal resolutions. We elaborate on this and several other examples in what follows.

- Cases may be lab-confirmed only or also suspected (with the definition of "suspected" varying across jurisdictions and time); they may be listed by date reported on the jurisdiction's website, by date reported to the public health authority, by date tested, by specimen collection date, or occasionally by symptom onset date (most informative but often unavailable or inapplicable). A casual review of many websites of local and state health departments suggests there is great heterogeneity in what is being reported [29].

- The term "hospitalizations" is used ambiguously, sometimes referring to incidence (hospital admissions) and sometimes to prevalence (hospital bed occupancy). These two quantities cannot easily be mapped to one another, because COVID-19 hospital discharges are rarely if ever reported. Furthermore, people admitted without a COVID-19 diagnosis may acquire the infection and/or the diagnosis at any time during their hospital stay.

- Hospitalizations may be reported by the location (typically county) of the patient's residence, but are more often reported by the county of the reporting hospital. This is an important distinction, as many rural COVID-19 patients in need of advanced care travel to the nearest secondary or tertiary hospital, often at a nearby urban county. For example, the hospitals in Pittsburgh, located in Allegheny County, Pennsylvania (population 1.2 million), treat patients from throughout a 13 county region in Southwestern Pennsylvania (population over 4 million). For tracking and forecasting hospitalization burden then, the geographic units of hospital referral regions (HRRs, [30]) may be most appropriate. Alas, these units do not conform to county boundaries, which complicates the projection of cases to hospitalizations.

- Deaths are usually reported by county where they occurred, which for hospitalized patients may differ from their county of residence [31].

- Hospitalization or deaths with COVID-19 are significantly different from hospitalization or deaths due to COVID-19 (as captured by e.g., a COVID-related chief complaint or primary ICD-10 code). The proportion of the two varies significantly by age groups and across time [32].
- Test positivity rates are most often reported by lumping together all tests performed regardless of the reason for performing them. Tests taken following positive diagnosis, due to symptoms, or due to being a contact of a confirmed case, are all likely to have a much higher positivity rate than that of the general population. Screening tests are most likely to reflect the true prevalence in the screened population. Sadly, very few jurisdictions report or even retain the breakdown of the test results by reason for testing, thereby losing forever valuable information.
- Medical insurance claims offer rich, detailed information about COVID-19 and other health conditions, but are not without their weaknesses. Claims are often not filed until weeks and months after the medical encounter. As such, signals derived from claims are usually subject to regular and considerable revisions up to 60 days after a given date of service, because signals must be updated each time new claims for that day of service are received (specific statistics on how this affects signal values are given below). This tends to make projections challenging, especially at finer geographic units such as counties, since there tends to be a high degree of heterogeneity across locations.
- Medical claims contain information about lab tests taken, but not their results. More generally, for understandable HIPAA reasons, medical claims contain only information necessary for adjudicating and auditing claims.

Data definitions must be disambiguated, clarified, and made consistent to the greatest extent possible, and remaining inconsistencies must be documented and saliently communicated.

Understanding the data generation process is critical for downstream applications. Both traditional public health surveillance data streams and newer digital surveillance streams are the result of often complex processes, some having to do with the underlying health status or activities being monitored, others with the reporting process itself. Understanding the entire "data generation process" for each data source can be challenging, but is absolutely essential for proper modeling and effective use of the data. Some examples are as follows.

- In medical claims, relevant diagnoses and comorbidities may not be reflected if they are not directly relevant to the charges incurred. On the other hand, because medical claim coding determines reimbursement levels, some codes may be over-represented relative to their medical significance.
- Some populations and some healthcare settings are not reflected in the commercial claims stream. These include the healthcare systems of the Department of Defense, Indian Health Services, Veterans Affairs, prison systems, and other systems that do not reimburse by procedure or service, as well as Medicare fee-for-service and Medicaid. This can cause significant bias in the signals relative to the prevalence in the general population.
- Public health reporting data are often subject to backlogs and reporting delays, and estimates for any particular date can be revised over time as errors are found or additional data becomes available. During the pandemic, audits, corrections, and the clearing of backlogs has frequently resulted in huge artificial spikes and drops [33]. Data aggregators like Johns Hopkins CSSE [34] have worked tirelessly to correct such anomalies after first publication (they attempt to back-distribute a spike or dip, by working with a local authority to figure out how this should best be done).
- Data revisioning (also known as "backfill") is pervasive not only in traditional public health reporting, but also in many (though not all) digital surveillance sources. As already alluded to above, signals based on medical claims typically undergo regular revisions because many claims (on which these signals are based) get submitted and processed late; for many COVID-related claims-based signals, the median relative error between initial reports and final values is over 10%, and only after 30 days or so do estimates typically match finalized values within 5% [25]. However, the systematic nature of these revisions suggests that, with suitable historical data, statistical models could be fit to estimate the final values from preliminary reports. By comparison, revisions to public health reports during the pandemic (the spikes and dips just described) have been much less systematic and much less predictable.

- Traditionally, public health agencies do not publish provisional data until it meets a level of stability. For example, NCHS data on the percentage of deaths due to pneumonia, influenza and COVID-19 is not released until at least 20% of the expected deaths in a jurisdiction have been reported [35], a process that may take several weeks. However, for modeling and forecasting purposes, even highly provisional data can be very informative, as long as sufficient historical provisional data is collected so that the statistical relationships between provisional values and finalized values can be modeled.

- Calendar effects permeate not only the reporting process but also health-seeking behavior and the epidemic process itself, with the effects on these three processes not easy to disentangle. Major holidays and other national or regional events are associated with significant travel, social mixing, and other distinct behaviors affecting disease transmission. However, holidays and weekends also affect health-seeking behavior via reduced non-emergency healthcare capacity (doctors' offices and labs being partially or fully closed). Perhaps the strongest calendar effects are on reporting, including claim filing and hospital reporting. Using 1- or 2-week trailing averages eliminates weekend effects, but at a cost of reduced temporal resolution, and it leaves unsolved the effects of holidays. A better approach might be to explicitly model and correct for calendar effects.

Mandated reporting in a time of emergency can be burdensome and inflexible. COVID-19 reporting by hospitals, as mandated by HHS during the pandemic, consisted of many dozens of data elements and imposed a significant burden on the nation's 6,000 or so hospitals at a time that they were already stretched to their limits. It also took a huge effort to formulate, communicate, disambiguate, and monitor for quality assurance and uniformity of interpretation. In light of this, it is not surprising that it took a long time, and pressure, for most hospitals to comply (near universal compliance was not achieved until December 2020). When changes needed to be made to the collected statistics, an arduous and time consuming process of approvals, re-formulation, re-communication, re-implementation, and re-assessment had to be followed.

While some aspects of mandated reporting are likely to remain irreplaceable, effective alternative surveillance sources can be of great use: they can improve on the timeliness, scope, robustness, and utility of mandated reporting data, while being less burdensome to collect. This is a common theme that runs through all four papers in the collection, but is perhaps most directly addressed in Reinhart et al. [25] (which focuses on the ecosystem of signals broadly). That said, we have far from saturated the utility of auxiliary surveillance. Much more needs to be developed in this area in order to usher epidemic tracking into its next phase of reliability, accuracy, and transparency. To us, electronic medical records hold a great promise for surveillance streams, and we elaborate on this in the next section.

Human behavior and its impact on the progression of epidemics is hard to measure and hard to model. In the nearly 10 years of government-organized epidemic forecasting exercises in the U.S., efforts were focused on modeling the natural history and likely evolution of the pathogen, with adaptation of human behavior playing a secondary role (if any role at all). The pandemic demonstrated that our forecasting models must pay closer attention to reactive human behavior, even more so if we are to consider interventions. Unfortunately, many highly relevant aspects of human behavior are not measured by publicly available data streams, like compliance with policies and recommendations (with perhaps surveys and mobility reports providing our best glimpse into these hard-to-observe aspects of behavior [36]), and even the very process of promulgating these policies and recommendations. Furthermore, even if we had these data in hand, incorporating their effects will require significant and new cognitive and behavioral modeling, with uncertain success. The tragic breakdown and fragmentation of trust in governments, public health officials, and healthcare professionals are perhaps the hardest factors to measure and model, yet they played an undeniable role in the progression of the current pandemic in the U.S. and other countries.

4 The Road Ahead

The focus of the Delphi group during the initial, critical period of the pandemic (February 2020 – March 2021) was on short-term goals: trying to provide signals, analysis, and decision support to federal, state and local public health officials, as well as to fellow researchers, data journalists, and the general public. In spring of 2021, equipped with the hard lessons learned during this tumultuous year, we turned our attention back to the original vision of the Delphi group, and asked ourselves: given where we are and what we know now, what is needed to be able to take a major step forward in epidemic tracking and forecasting? In this section, we list some ideas which we hope will elicit further public discussion and, most importantly, experimentation. Because our expertise lies in modeling and forecasting, not in public health surveillance, our perspective and recommendations are necessarily limited to those aspects of surveillance that are needed to realize our vision.

EMR as a key missing component for epidemic tracking and forecasting. The success of nowcasting, analytics and forecasting depends crucially on the availability of rich, real-time data sources. In light of the limitations of mandated reporting discussed above, we must consider the complementary value of other data sources. Chief among these are electronic medical records (EMR), as are being created and used daily by inpatient and outpatient healthcare providers, medical laboratories, and pharmacies. The advantage of these data resources: they are rich, real-time, and already being generated (found "in the wild"). The challenges: they are highly fragmented in the U.S., with its approximately 6,000 hospitals and 100,000 outpatient care facilities. One promising avenue for countering this fragmentation are Health Information Exchanges (HIEs), which were set up in the early 2000s with the support of the federal government to coordinate the sharing of healthcare information among healthcare providers in a given region, and eventually nationally. The primary goal of the HIEs has been patient continuity of care, but public health surveillance is recognized as an additional worthy goal. In the context of health surveillance, HIEs hold the promise of reducing fragmentation from a hundred thousand partners to only a few hundred.

Other formidable challenges to using EMR are legal, ethical, commercial, and operational. Who owns the data, who has access to it, and who has use rights are all complex and often open questions. An overriding concern is of course patient privacy. We must find a way to use this highly promising data for the common good without compromising the privacy of individuals. Fortunately, a technological solution appears feasible, in the form of *federated surveillance*. An outline is as follows.

- A common API is developed for querying all participating EMR custodians.
- EMR databases are queried daily with an agreed upon set of queries, and return aggregated counts.
- These counts are further aggregated across multiple providers and localities, and then fused with all other available data sources to provide alerts, nowcasts, and forecasts.
- These model outputs are then shared back to the contributing EMR custodians, as well as to CDC and other federal and state agencies.
- Done in this way, no personal health information (PHI) ever leaves the custodians' premises, while aggregated statistics can be combined to increase statistical power, thereby shortening alert latency and improving detection and prediction capabilities.

A successful example of federated querying (albeit designed for research rather than real-time tracking and forecasting) has been recently demonstrated in the UK [37].

One advantage of this approach to health surveillance is that when a new emerging health crisis is identified, or when specific syndromic signatures are discovered (e.g., ageusia and anosmia for COVID-19), a new query can be developed, approved, and deployed literally overnight, allowing us to "shine a light" on it on very short notice. This can be contrasted with traditional, legally-mandated public health reporting, which could take weeks and months to develop, approve, negotiate, disseminate, implement, monitor, and assure quality of, as has happened during the current pandemic. In the slightly longer term, demonstrating the effectiveness and superiority of federated surveillance could obviate the need for crisis-time mandated reporting, alleviating the reporting burden on hospitals and other healthcare providers during these difficult times.

In an interpandemic period, an important use case for federated surveillance is detection of trends and anomalies. A set of queries can be designed to continuously test for unusual recent spikes or trends in any number of diagnoses, symptoms, lab tests, or prescriptions. The aggregation of evidence across many systems, localities, and data streams will make detection both more sensitive and more robust. Such a system would likely have detected the opioid epidemic years before it was actually noticed.

One open technical challenge with the federated surveillance approach is semantic heterogeneity: the use of emerging Health IT standards like HL7's FHIR [38] can enable a unified view of EMR data elements, but different healthcare systems often have different operational definitions for supposedly universal concepts like "high blood pressure" or "low blood oxygenation". Combining counts of such events across different healthcare systems may be a bit like adding apples to oranges. It will take some work to harmonize semantics across so many diverse data custodians, but this is both doable and well worth doing. Note that this problem is less severe for the many surveillance and anomaly detection tasks where the focus is on changes in a signal (in a given location) over time, rather than on its absolute meaning.

Different phases of epidemic surveillance call for different analytic tools. It is important to discuss analytic needs separately for each of three different phases of epidemic surveillance and tracking, since each poses different technical challenges and requires different analytical tools.

In the *interpandemic phase*, the main activity is threat scanning for threats, namely monitoring data streams and events throughout the world for disconcerting developments. Relevant statistical tools include anomaly detection and scan statistics to help decide when an epidemiological investigation is warranted. While it may be possible to rank the risks of different outbreak triggering events (like species jumping or point mutations) in different locations, which could in turn be used to inform surveillance resource allocation, conventional forecasting has a limited role to play in this phase, as such events have large inherent uncertainty.

In the *containment phase*, a discovered threat must be intensely monitored, continuously assessed, and ultimately contained. The analytical, data-driven tools required in this phase include real-time estimation of critical epidemiological parameters such as R0, infection fatality rates, the incubation period, the serial interval, and so on. These real-time estimates are necessarily based on provisional data, highlighting the value of modeling the data generation process discussed above. In this phase, forecasting still has a limited role, since the outbreak is still local, its fundamental dynamics unclear, and point events can have large consequences down the road.

If containment fails, during the *mitigation phase* the goals of analytics expand significantly to include informing mitigation policies and planning. Real-time tracking (nowcasting) and short term-forecasting (a few weeks ahead) can play critical roles in these activities, and indeed have been the focus of our group's work since its inception. While there is still important work to be done and advances to be made in this area, we believe that these advances are likely to be incremental until we see major progress in either (1) supporting data streams (e.g., better standardization and cleaning of public health reporting data, identification of leading indicators from, say, EMR); or (2) our collective scientific understanding of the real-world geotemporal dynamics of epidemics (discussed next).

Useful, reliable longer-term forecasting remains an aspiration. Influenza forecasting exercises in the last several years demonstrated that it is often possible to usefully quantify uncertainty over the remainder of an ongoing flu season [3]. But this success was based mostly on observing the behavior of seasonable epidemics over several decades. To reliably forecast the progression of pandemics, where relevant historical data is almost nonexistent, we must have a detailed quantitative understanding of how different, diverse factors affect disease transmissibility. Such an understanding is currently grossly lacking, as evidenced by our collective failure to predict [39] (or even understand, post-hoc [40]) the high-level temporal and geographic contours of the main pandemic waves in the U.S. Yet this very pandemic, the most instrumented in human history, is also a rare opportunity to attempt this vital scientific and technological goal.

Acknowledgements

Authors' note: to be filled out later.

References

[1] Matthew Biggerstaff, David Alper, Mark Dredze, Spencer Fox, Isaac Chun-Hai Fung, Kyle S Hickmann, Bryan Lewis, Roni Rosenfeld, Jeffrey Shaman, Ming-Hsiang Tou, Paola Velardi, Alessandro Vespignani, Lyn Finelli, and the Influenza Forecasting Contest Working Group. Results from the centers for disease control and prevention's predict the 2013–2014 Influenza Season Challenge. *BMC Infectious Diseases*, 16(1):357, 2016.

[2] Matthew Biggerstaff, Michael Johansson, David Alper, Logan C Brooks, Prithwish Chakraborty, David C Farrow, Sangwon Hyun, Sasikiran Kandula, Craig McGowan, Naren Ramakrishnan, Roni Rosenfeld, Jeffrey Shaman, Robert Tibshirani, Ryan J Tibshirani, Alessandro Vespignani, Wan Yang, Qian Zhang, and Carrie Reed. Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics*, 24: 26–33, 2018.

[3] Nicholas G Reich, Logan C Brooks, Spencer J Fox, Sasikiran Kandula, Craig J McGowan, Evan Moore, Dave Osthus, Evan L Ray, Abhinav Tushar, Teresa K Yamana, Matthew Biggerstaff, Michael A Johansson, Roni Rosenfeld, and Jeffrey Shaman. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proceedings of the National Academies of Sciences*, 116(8):3146–3154, 2019.

[4] Michael A Johansson, Karyn M Apfeldorf, Scott Dobson, Jason Devita, Anna L Buczak, Benjamin Baugher, Linda J Moniz, Thomas Bagley, Steven M Babin, Erhan Guven, Teresa K Yamana, Jeffrey Shaman, Terry Moschou, Nick Lothian, Aaron Lane, Grant Osborne, Gao Jiang, Logan C Brooks, David C Farrow, Sangwon Hyun, Ryan J Tibshirani, Roni Rosenfeld, Justin Lessler, Nicholas G Reich, Derek A T Cummings, Stephen A Lauer, Sean M Moore, Hannah E Clapham, Rachel Lowe, Trevor C Bailey, Markel García-Díez, Marilia Sá Carvalho, Xavier Rodó, Tridip Sardar, Richard Paul, Evan L Ray, Krzysztof Sakrejda, Alexandria C Brown, Xi Meng, Osonde Osoba, Raffaele Vardavas, David Manheim, Melinda Moore, Dhananjai M Rao, Travis C Porco, Sarah Ackley, Fengchen Liu, Lee Worden, Matteo Convertino, Yang Liu, Abraham Reddy, Eloy Ortiz, Jorge Rivero, Humberto Brito, Alicia Juarrero, Leah R Johnson, Robert B Gramacy, Jeremy M Cohen, Erin A Mordecai, Courtney C Murdock, Jason R Rohr, Sadie J Ryan, Anna M Stewart-Ibarra, Daniel P Weikel, Antarpreet Jutla, Rakibul Khan, Marissa Poultney, Rita R Colwell, Brenda Rivera-García, Christopher M Barker, Jesse E Bell, Matthew Biggerstaff, David Swerdlow, Luis Mier-Y-Teran-Romero, Brett M Forshey, Juli Trtanj, Jason Asher, Matt Clay, Harold S Margolis, Andrew M Hebbeler, Dylan George, and Jean-Paul Chretien. An open challenge to advance probabilistic forecasting for dengue epidemics. *Proceedings of the National Academy of Sciences*, 116(48):24268–24274, 2019.

[5] Sara Y Del Valle, Benjamin H McMahon, Jason Asher, Richard Hatchett, Joceline C Lega, Heidi E Brown, Mark E Leany, Yannis Pantazis, David J Roberts, Sean Moore, A Townsend Peterson, Luis E Escobar, Huijie Qiao, Nicholas W Hengartner, and Harshini Mukundan. Summary results of the 2014-2015 DARPA Chikungunya challenge. *BMC Infectious Diseases*, 18(245), 2018.

[6] Marco Ajelli, Qian Zhang, Kaiyuan Sun, Stefano Merler, Laura Fumanelli, Gerardo Chowell, Lone Simonsen, Cecile Viboud, and Alessandro Vespignani. The RAPIDD Ebola forecasting challenge: Model description and synthetic data generation. *Epidemics*, 22:3–12, 2018.

[7] Cécile Viboud, Kaiyuan Sun, Robert Gaffey, Marco Ajelli, Laura Fumanelli, Stefano Merler, Qian Zhang, Gerardo Chowell, Lone Simonsen, Alessandro Vespignani, and RAPIDD Ebola Forecasting Challenge Group. The RAPIDD Ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics*, 22:13–21, 2018.

[8] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.

[9] John S Brownstein, Clark C Freifeld, and Lawrence C Madoff. Digital disease detection — harnessing the web for public health surveillance. *New England Journal of Medicine*, 360(21):2153–2157, 2009.

[10] Marcel Salathé, Linus Bengtsson, Todd J Bodnar, Devon D Brewer, John S Brownstein, Caroline Buckee, Ellsworth M Campbell, Ciro Cattuto, Shashank Khandelwal, Patricia L Mabry, and Alessandro Vespignani. Digital epidemiology. *PLOS Computational Biology*, 8(7):1–3, 2012.

[11] Taha A Kass-Hout and Hend Alhinnawi. Social media in public health. *British Medical Bulletin*, 108(1):5–24, 2013.

[12] Mauricio Santillana, André T Nguyen, Mark Dredze, Michael J Paul, Elaine O Nsoesie, and John S Brownstein. Combining search, social media, and traditional data sources to improve influenza surveillance. *PLOS Computational Biology*, 11(10):e1004513, 2015.

[13] Michael J Paul and Mark Dredze. Social monitoring for public health. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 9(5):1–183, 2017.

[14] Taha A Kass-Hout and Xiaohui Zhang. *Biosurveillance: Methods and Case Studies*. CRC Press, 2011.

[15] Sandra J Carlson, Craig B Dalton, Michelle T Butler, John Fejsa, Elissa Elvidge, and David N Durrheim. Flutracking weekly online community survey of influenza-like illness annual report 2011 and 2012. *Communicable diseases intelligence quarterly report*, 37(4):E398–406, 2013.

[16] Cécile Viboud, Vivek Charu, Donald Olson, Sébastien Ballesteros, Julia Gog, Farid Khan, Bryan Grenfell, and Lone Simonsen. Demonstrating the use of high-volume electronic medical claims data to monitor local and regional influenza activity in the us. *PLOS ONE*, 9(7):1–12, 07 2014. doi: 10.1371/journal.pone.0102429. URL https://doi.org/10.1371/journal.pone.0102429.

[17] Mark S. Smolinski, Adam W. Crawley, Kristin Baltrusaitis, Rumi Chunara, Jennifer M. Olsen, Oktawia Wójcik, Mauricio Santillana, Andre Nguyen, and John S. Brownstein. Flu Near You: Crowdsourced symptom reporting spanning 2 influenza seasons. *American Journal of Public Health*, 105(10):2124–2130, 2015. doi: 10.2105/AJPH.2015.302696. URL https://doi.org/10.2105/AJPH.2015.302696. PMID: 26270299.

[18] Mauricio Santillana, Andre T Nguyen, Tamara Louie, Anna Zink, Josh Gray, Iyue Sung, and John S Brownstein. Cloud-based electronic health records for real-time, region-specific influenza surveillance. *Scientific Reports*, 6(1): 1–8, 2016.

[19] Vivek Charu, Scott Zeger, Julia Gog, Ottar N. Bjørnstad, Stephen Kissler, Lone Simonsen, Bryan T. Grenfell, and Cécile Viboud. Human mobility and the spatial transmission of influenza in the united states. *PLOS Computational Biology*, 13(2):1–23, 02 2017. doi: 10.1371/journal.pcbi.1005382. URL https://doi.org/10.1371/journal.pcbi.1005382.

[20] Carl E Koppeschaar, Vittoria Colizza, Caroline Guerrisi, Clément Turbelin, Jim Duggan, W John Edmunds, Charlotte Kjelsø, Ricardo Mexia, Yamir Moreno, Sandro Meloni, Daniela Paolotti, Daniela Perrotta, Edward van Straten, and Ana O Franco. Influenzanet: Citizens among 10 countries collaborating to monitor influenza in europe. *JMIR Public Health Surveill*, 3(3):e66, Sep 2017. ISSN 2369-2960. doi: 10.2196/publichealth.7429. URL http://publichealth.jmir.org/2017/3/e66/.

[21] Cheng-Yi Yang, Ray-Jade Chen, Wan-Lin Chou, Yuarn-Jang Lee, and Yu-Sheng Lo. An integrated influenza surveillance framework based on national influenza-like illness incidence and multiple hospital electronic medical records for early prediction of influenza epidemics: Design and evaluation. *Journal of Medical Internet Research*, 21 (2):e12341, Feb 2019. ISSN 1438-8871. doi: 10.2196/12341. URL http://www.jmir.org/2019/2/e12341/.

[22] Sequoia I. Leuba, Reza Yaesoubi, Marina Antillon, Ted Cohen, and Christoph Zimmer. Tracking and predicting U.S. influenza activity with a real-time surveillance network. *PLOS Computational Biology*, 16(11):1–14, 11 2020. doi: 10.1371/journal.pcbi.1008180. URL https://doi.org/10.1371/journal.pcbi.1008180.

[23] Jennifer M Radin, Nathan E Wineinger, Eric J Topol, and Steven R Steinhubl. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: A population-based study. *The Lancet Digital Health*, 2(2):e85–e93, 2020.

[24] Sarah F Ackley, Sarah Pilewski, Vladimir S Petrovic, Lee Worden, Erin Murray, and Travis C Porco. Assessing the utility of a smart thermometer and mobile application as a surveillance tool for influenza and influenza-like illness. *Health Informatics Journal*, 26(3):2148–2158, 2020. doi: 10.1177/1460458219897152. URL https://doi.org/10.1177/1460458219897152. PMID: 31969046.

[25] Alex Reinhart, Logan Brooks, Maria Jahja, Aaron Rumack, Jingjing Tang, , Sumit Agrawal, Wael Al Saeed, Taylor Arnold, Amartya Basu, Jacob Bien, Ángel A. Cabrera, Andrew Chin, Eu Jing Chua, Brian Clark, Sarah Colquhoun, Nat DeFries, David C. Farrow, Jodi Forlizzi, Jed Grabman, Samuel Gratzl, Alden Green, George Haff,

Robin Han, Kate Harwood, Addison J. Hu, Raphael Hyde, Sangwon Hyun, Ananya Joshi, Jimi Kim, Andrew Kuznetsov, Wichada La Motte-Kerr, Yeon Jin Lee, Kenneth Lee, Zachary C. Lipton, Michael X. Liu, Lester Mackey, Kathryn Mazaitis, Daniel J. McDonald, Phillip McGuinness, Balasubramanian Narasimhan, Michael P. O'Brien, Natalia L. Oliveira, Pratik Patil, Adam Perer, Collin A. Politsch, Samyak Rajanala, Dawn Rucker, Chris Scott, Nigam H. Shah, Vishnu Shankar, James Sharpnack, Dmitry Shemetov, Noah Simon, Benjamin Y. Smith, Vishakha Srivastava, Shuyi Tan, Robert Tibshirani, Elena Tuzhilina, Ana Karina Van Nortwick, Valérie Ventura, Larry Wasserman, Benjamin Weaver, Jeremy C. Weiss, Spencer Whitman, Kristin Williams, Roni Rosenfeld, and Ryan J. Tibshirani. An open repository of real-time covid-19 indicators. medRxiv, 2021.

[26] Daniel J. McDonald, Jacob Bien, Alden Green, Addison J. Hu, Nat DeFries, Sangwon Hyun, Natalia L. Oliveira, James Sharpnack, Jingjing Tang, Robert Tibshirani, Valérie Ventura, Larry Wasserman, and Ryan J. Tibshirani. Can auxiliary indicators improve COVID-19 forecasting and hotspot prediction? medRxiv, 2021.

[27] Reich Lab. The COVID-19 Forecast Hub. https://covid19forecasthub.org, 2020.

[28] Joshua A. Salomon, Alex Reinhart, Alyssa Bilinski, Eu Jing Chua, Wichada La Motte-Kerr, Minttu M. Rönn, Marissa Reitsma, Katherine Ann Morris, Sarah LaRocca, Tamer Farag, Frauke Kreuter, Roni Rosenfeld, and Ryan J. Tibshirani. The U.S. COVID-19 Trends and Impact Survey, 2020-2021: Continuous real-time measurement of COVID-19 symptoms, risks, protective behaviors, testing and vaccination. medRxiv, 2021.

[29] Sara Simon. Inconsistent reporting practices hampered our ability to analyze COVID-19 data. Here are three common problems we identified. COVID Tracking Project, https://covidtracking.com/analysis-updates/three-covid-19-data-problems, 2021.

[30] John E. Wennberg and Megan McAndrew Cooper. *The Dartmouth Atlas of Health Care in the United States*. American Hospital Publishing, Chicago, 1998.

[31] National Center for Health Statistics. Provisional death counts for coronavirus disease 2019 (COVID-19). https://www.cdc.gov/nchs/nvss/vsrr/COVID19/index.htm, 2021.

[32] Nathanael Fillmore, Jennifer La, Chunlei Zheng, Shira Doron, Nhan Do, Paul Monarch, and Westyn Branch-Elliman. The COVID-19 hospitalization metric in the pre- and post-vaccination eras as a measure of pandemic severity: A retrospective, nationwide cohort study. Research Square, 2021.

[33] Simone Arvisais-Anhalt, Christoph U Lehmann, Jason Y Park, Ellen Araj, Michael Holcomb, Andrew R Jamieson, Samuel McDonald, Richard J Medford, Trish M Perl, Seth M Toomay, Amy E Hughes, Melissa L McPheeters, and Mujeeb Basit. What the coronavirus disease 2019 (COVID-19) pandemic has reinforced: The need for accurate data. *Clinical Infectious Diseases*, 72(6):920–923, 11 2021. ISSN 1058-4838.

[34] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5):533–534, 2020.

[35] National Center for Health Statistics. Pneumonia and influenza mortality surveillance from the national center for health statistics mortality surveillance system. https://gis.cdc.gov/grasp/fluview/mortality.html, 2021.

[36] Alyssa Bilinski, Ezekial Emanuel, Joshua A Salomon, and Atheendar Venkataramani. Better late than never: Trends in COVID-19 infection rates, risk perceptions, and behavioral responses in the USA. *Journal of General Internal Medicine*, 36(6):1825–1828, 2021.

[37] Elizabeth J Williamson, Alex J Walker, Krishnan Bhaskaran, Seb Bacon, Chris Bates, Caroline E Morton, Helen J Curtis, Amir Mehrkar, David Evans, Peter Inglesby, Jonathan Cockburn, Helen I McDonald, Brian MacKenna, Laurie Tomlinson, Ian J Douglas, Christopher T Rentsch, Rohini Mathur, Angel YS Wong, Richard Grieve, David Harrison, Harriet Forbes, Anna Schultze, Richard Croker, John Parry, Frank Hester, Sam Harper, Rafael Perera, Stephen JW Evans, Liam Smeeth, , and Ben Goldacre. OpenSAFELY: Factors associated with COVID-19 death in 17 million patients. *Nature*, 584:430–436, 2020.

[38] HL7 Community. Welcome to FHIR. https://www.hl7.org/fhir/, 2021.

[39] Nicholas G Reich, Ryan J Tibshirani, Evan L Ray, and Roni Rosenfeld. On the predictability of COVID-19. https://forecasters.org/blog/2021/09/28/on-the-predictability-of-covid-19/, 2021. International Institute of Forecasters Blog.

[40] Hawre Jalal, Kyueun Lee, and Donald S Burke. Prominent spatiotemporal waves of COVID-19 incidence in the United States: Implications for causality, forecasting, and control. medRxiv, 2021.