Chapter 15 Digital Epidemiology



Yelena Mejova

Abstract Computational social science has had a profound impact on the study of health and disease, mainly by providing new data sources for all of the primary Ws-what, who, when, and where-in order to understand the final "why" of disease. Anonymized digital trace data bring a new level of detail to contact networks, search engine and social media logs allow for the now-casting of symptoms and behaviours, and media sharing informs the formation of attitudes pivotal in health decision-making. Advances in computational methods in network analysis, agentbased modelling, as well as natural language processing, data mining, and time series analysis allow both the extraction of fine-grained insights and the construction of abstractions over the new data sources. Meanwhile, numerous challenges around bias, privacy, and ethics are being negotiated between data providers, academia, the public, and policymakers in order to ensure the legitimacy of the resulting insights and their responsible incorporation into the public health decision-making. This chapter outlines the latest research on the application of computational social science to epidemiology and the data sources and computational methods involved and spotlights ongoing efforts to address the challenges in its integration into policymaking.

15.1 Introduction

From the beginnings of epidemiology, the importance of data has been central. Often considered fathers of the field, John Graunt analysed London's bills of mortality to measure the mortality of certain diseases in 1663, and later in 1854, John Snow mapped the cholera cases to identify its sources. Although since those early days in London the medical and mathematical understanding of disease have greatly advanced, one of the primary roles of the epidemiologist is still to prepare and

Y. Mejova (🖂)

ISI Foundation, Turin, Italy e-mail: yelenamejova@acm.org

[©] The Author(s) 2023 E. Bertoni et al. (eds.), *Handbook of Computational Social Science for Policy*, https://doi.org/10.1007/978-3-031-16624-2_15

organize the collection of relevant and useful data and to use it to model disease (Obi et al., 2020). This data includes the fundamental *Ws* that are necessary to understand disease: health event (what), people involved (who), place (where), time (when), and causes, risk factors, and modes of transmission (why/how) (Dicker et al., 2006). Thus, some of the main tasks of an epidemiologist are disease surveillance, field investigation, contact tracing, evaluation of interventions, and public communication—all of which have been transformed by the digital and computing revolutions.

Scientifically, the field is highly multidisciplinary, first measuring the basics of the Ws-identifying the people, places, and time frames of the health eventsand then introducing higher-order considerations, the biology of disease, behaviour of its carriers, and ecological influences on the transmission. By building models around this knowledge, it attempts to recommend possible interventions, which then require additional measurement and modelling of complex feedback effects and the psychological and behavioural factors. Advances in disparate fields like genetics, behavioural economics, and ecology on the one hand and more recent strides in computing methods and digitization on the other are making it possible for epidemiology to develop a systems conceptualization of the fields it connects. Computational social science (CSS) in particular adds new tools via large-scale detection, tracking, and contextualizing of disease. As we will see below, digital traces such as mobility and cellphone data have been used to better understand human networks, user-generated content on social media and the web has been employed to now-cast symptoms and disease, and social interactions have been monitored to understand the impact of social contact and new information on healthrelated behaviour change. Capturing the latest modelling and computing techniques, the umbrella terms of *digital* or *computational epidemiology* encompass these new methodological developments.

A string of epidemics in the early twenty-first century-H1N1 (swine flu) in 2009, Ebola in 2014 and 2019, and Zika in 2016-has brought epidemiology to the forefront of public awareness, culminating in the COVID-19 pandemic (at the time of this writing, in any case). Meanwhile, public health policy and interventions are being increasingly informed by telecommunications and other digital data (Budd et al., 2020; Oliver et al., 2020; Rich & Miah, 2017). Governments are collaborating with major cellphone companies to perform privacy-preserving contact tracing, internet companies are releasing aggregated mobility data for contagion modelling, and social media giants are partnering with public health organizations to tackle health misinformation and to support public health messaging campaigns. Throughout, a constant negotiation is at play between the needs of public health researchers and the release of commercially valuable information by the companies. Moreover, a less publicized, but nevertheless critical, battle is being waged against non-communicable diseases including cardiovascular diseases, cancer, diabetes, and mental health disorders. Daily digital traces, such as social media posts and location check-ins, are being used to understand the lifestyle choices of large cohorts, as an alternative to surveys and diaries. Discussions around mental health, disordered eating, illicit drugs, and other topics that are difficult to capture using traditional surveillance methods are now presenting a window into vulnerable populations, even before they register in medical records.

Despite the great promise of new data sources and methodologies, big data approaches are a subject to a slew of challenges that the field needs to overcome in order to establish fruitful collaboration with policymakers. Although big, the datasets often present biased view of the population, which is more tech-savvy and affluent, while excluding those who may have a more urgent need of monitoring and assistance. However, integration of this new data into existing datasets allows for the reduction of overall bias and helps in extending analyses performed on traditional data sources. Encompassing many disciplines who have their own organization, research frameworks, peculiar jargon, and publication venues, digital epidemiology is still in the processing of bridging the siloes to encourage truly multidisciplinary insight. Standardizing the reporting and transparency among these disciplines aims to reduce the number of isolated studies which may suffer from the lack of reproducibility due to the peculiar nature of the available data, application domain, or poorly documented methodology. The legal and ethical standards of using digital data are still being decided through a dialogue between data owners, public health researchers, academics in various disciplines, and representatives of the users of the digital platforms. Thus, the field is still building the structures of cooperation, trust, and legitimacy that are necessary to provide impactful insights for policymakers. Nevertheless, COVID-19 has accelerated the integration of digital epidemiology into its decision-making process. Below, we outline the major accomplishments in the application of computational social science to epidemiology, the accompanying challenges, and the possible ways forward to greater legitimacy and impact.

15.2 Existing Literature

The explosion in the utilization of computational methods for epidemiology has been spurred by the combination of new computational techniques and the availability of new sources of data. The immense volume of available data has encouraged further development and integration into the scientific toolkit of distributed computing frameworks, as well as data-intense deep learning algorithms, with frameworks such as Apache Spark and TensorFlow that allow the ingestion and processing of terabytes of data (Kleppmann, 2017; Weidman, 2019). The rise of infrastructure as a service (IaaS) business model from giants of industry including Amazon Web Services, Microsoft Azure, and cloud services from Oracle, Google, and IBM has allowed the researchers to access sophisticated infrastructure without purchasing the hardware and support staff within their institutions (on a similar topic, see also Fontana & Guerzoni, 2023).

Much of the data that has accompanied these developments in the computing field has been put to use by epidemiologists, opening new scientific ground. The ongoing digitization of medical records, insurance claims, and governmental public health data continues to provide large-scale, high-quality view of individuals within the medical system. Ongoing efforts, such as the European Health Data Space,¹ aggregate such datasets, handle privacy concerns, and make it available for research and policymaking (European Commission., 2021). Moreover, the communication revolution has enabled researchers to better understand these individuals even before they enter the public health system. Digital traces of people's daily activities, including the apps they use, web searches they make, social media posts they publish, as well as the signals from the wearables they keep on their bodies, can help create a view of health-related activities with an unprecedented resolution and reach. One of the earliest attempts to track influenza-like illness (ILI) using user-generated data was proposed by a team of Google researchers who tracked the occurrence of specific keywords in the company's search query logs (Ginsberg et al., 2009). Although highly criticized by subsequent researchers (Lazer et al., 2014) (we will discuss these concerns below), research on web logs continues to produce encouraging results, including detecting adverse reactions (Yom-Tov & Lev-Ran, 2017), predicting diagnosis of diabetes (Hochberg et al., 2019), and understanding the information needs around medical topics (Rosenblum & Yom-Toy, 2017). Specialized application use has been used to understand the effects of gamification (Althoff et al., 2016) and social contagion (Aral & Nicolaides, 2017) on exercise and the characteristics of (un-)successful diets (Weber & Achananuparp, 2016). The text posted by thousands of users on social media platforms has been used to identify and track depression (De Choudhury et al., 2013), eating disorders (Stewart et al., 2017), attitudes toward vaccination (Cossard et al., 2020), and other health interventions. The networked nature of the data often allows the study of the way in which information (Johnson et al., 2020), behaviours, and diseases propagate. Finally, anonymized mobility data, often coming from telephone and transportation companies, has allowed a more fine-grained transmission modelling of the disease (Vespe et al., 2021), as well as the impact of mobility-related interventions (Jeffrey et al., 2020). These data sources add immense value to the traditional ones by increasing the population coverage (some into millions of people), temporal resolution (allowing "now-casting"), and qualitative depth that are impossible or prohibitively expensive to reach outside the digital domain.

One of the earliest examples of the application of computational models to infectious diseases was human influenza, which is an ongoing public health battle. It is continuously analysed via viral phylodynamics in order to better understand its transmission dynamics. Computational phylogenetics methods are applied to datasets of genetic sequences sampled over time and sub-populations in order to assemble a phylogenetic tree and estimate various dynamics of the process (Volz et al., 2013). Fitness models also help in selecting the vaccines year over year (Łuksza & Lässig, 2014). Beyond the study of the virus itself, CSS has introduced several behavioural aspects to the models, many of which have been used during the COVID-19 epidemic. Mobility data (including that provided publicly by large corporations during the pandemic) has been used to monitor the compliance with

¹ https://ec.europa.eu/health/ehealth-digital-health-and-care/european-health-data-space_en

interventions, such as the stay-at-home orders during COVID-19, revealing the role of awareness and fatigue in modelling risky behaviours (Weitz et al., 2020). Largescale online surveys and crowdsourcing have been used to gauge psychological and behavioural responses to the pandemic around the world (Yamada et al., 2021). Even larger efforts, such as InfluenzaNet, recruit thousands of volunteers across Europe to regularly report ILI symptoms, allowing researchers to identify risk factors and gauge influenza vaccine effectiveness (Koppeschaar et al., 2017). Travel records have been used to track the international transmission of disease (Azad & Devi, 2020), whereas a machine-learned anonymized smartphone mobility map has been used to forecast influenza within and across countries (Venkatramanan et al., 2021). For instance, the Global Epidemic and Mobility (GLEaM) framework uses local and international mobility data to build epidemic models, allowing for the simulation of worldwide pandemics, including estimating the impact of interventions during the COVID-19 epidemic (Chinazzi et al., 2020; Van den Broeck et al., 2011). To better understand the reasons behind risky behaviours and non-compliance with public health advice, researchers utilized discussions on social media, often finding misunderstandings and downright misinformation (Betti et al., 2021; Keller et al., 2021). Finally, public health communication campaigns have been evaluated using outreach online by influencers (Bonnevie et al., 2020) as well as news websites and popular social media sites (Carlson et al., 2020).

Unlike in the beginning of epidemiology's development as a science, the infectious diseases have these days given way to non-communicable diseases as the cause of illness and death, especially in the developed countries. The daily behaviours captured in digital trace data, especially social media, have been extensively used to study non-communicable diseases including obesity and diabetes type 2, mental illness, and even suicide. At population level, diabetes has been tracked using store purchase data (Aiello et al., 2019), as well as social media posts (Abbar et al., 2015), and some environmental causes have been tracked in the USA, with a focus on "food deserts" where access to healthy food is limited (De Choudhury, Sharma, et al., 2016). Attempts to inform potential interventions have been made by measuring the importance of community support during a weight loss journey (Cunha et al., 2016) and the effect of intervention messaging on those affected by anorexia (Yom-Tov et al., 2012). Observational studies of exercise in particular through specialized exercise applications have shown that information about other people's routine may affect one's own (Aral & Nicolaides, 2017) and that gender plays an important role in the continued use of such apps (Mejova & Kalimeri, 2019). Further, a combination of web search and wearables data has been used to show the health impact of applications not necessarily meant for exercise, such as Pokémon Go, which resulted in potentially years worth of life spans added to the fans of the game (Althoff et al., 2016). The anonymous and connected nature of social media and specialized forums have also allowed a better understanding of depression, anxiety, eating disorders, and other mental health issues (for an overview, see Chancellor & De Choudhury, 2020). The text of the posts has been used to predict suicidal ideation (Cheng et al., 2017), psychotic relapses (Birnbaum et al., 2019), and PTSD (Coppersmith et al., 2014). More specialized data sources have been used to track recreational drug use (Deluca et al., 2012), as well as the use of "dark web" as a marketplace for such activities (Aldridge & Décary-Hétu, 2016). In combination with screening questionnaires which use validated scales such as Center for Epidemiologic Studies Depression Scale (CES-D) and Beck Depression Inventory (BDI), the daily self-expressions of those dealing with mental health issues provide an unintrusive record of the condition's progression and reactions to potential interventions.

These encouraging developments have been accompanied by a vigorous discussion of their limitations. The privacy concerns regarding secondary use of personal data, even if originally posted on public platforms, demand a critical evaluation of the balance between potential benefits of public health research, compared to the privacy risks to the individuals captured in the data (see, e.g. Taylor, 2023). Other critiques are more unique to the field of epidemiology. For instance, the machine learning framework of classification, as well as most deterministic compartmental models (such as Susceptible-Infected-Recovered (SIR), more on which later), makes necessary simplifying assumptions about the natural progression of a disease, its behaviour, as well as the pharmaceutical and non-pharmaceutical interventions introduced to slow its spread, although more sophisticated models with more complex representations are continuously being proposed.

The separation between traditional epidemiology and computing disciplines in the research teams often results in the failure to take into consideration the established theories in clinical science, using operationalization that is most convenient technically, but not as well matched to the medical condition tracked, while a vague communication of the technical aspects of computing pipelines makes it difficult to integrate the results into clinical practice (Chancellor & De Choudhury, 2020). Observational studies have also lacked the rigor of causal analysis, often stopping at correlational observations. Despite capturing multitudes of people, each data source has substantial biases that must be not only acknowledged by the researchers but accounted for in the analytical pipeline (Yom-Tov, 2019). Finally, data ownership, global justice, and ethical oversight are all important problems that need to be addressed for digital epidemiology to gain legitimacy on the scientific and policy stage (Vayena et al., 2015). We will touch on these and other peculiarities of using computational social science for epidemiology in the next section.

15.3 Computational Guidelines

The abovementioned literature not only pushes the boundaries of traditional epidemiology and the purview of computing but addresses multiple important policy questions regarding public health. The third goal of the UN Sustainable Development Goals (SDGs) is to "Ensure healthy lives and promote well-being for all at all

ages".² For instance, the goal encompasses the work on alleviating communicable and non-communicable diseases, prevention and treatment of substance abuse, ensuring access to sexual and reproductive services, and increasing the healthcare capacity in all countries, but especially in the developing ones. Although CSS cannot build the necessary infrastructure, it can measure, on both community and individual scale, the utilization of healthcare services, the barriers experienced by the populous, and the expression of unfulfilled needs. Furthermore, it can help in tracking and forecasting disease, again at the scales including individuals, thus measuring the impact of potential ongoing interventions. In fact, CSS can help to craft, deploy, and monitor epidemiological interventions by providing detailed profiling of the target audience, individualized message delivery, and fine-grained behavioural feedback. In order to bring these promises to fruition, a slew of challenges remain to be fully addressed by the research and policy community, including data access and privacy, construct validity, methodological transparency, sampling bias, accounting for confounders, and finally sufficiently clear communication to ensure real-world application. Below, we discuss several policy questions that CSS may address and outline technical and organizational best practices.

15.3.1 Infectious Diseases

The modelling and predicting of infectious diseases is perhaps the most well-known purview of digital epidemiology. Some of the simplest models of disease spread use a system of states as a basis, such as the Susceptible-Infected-Recovered (SIR) model wherein the population can be put into one of these three states (Bjørnstad et al., 2020). Other compartmental models exist which describe the progression of disease with more states ("compartments"), including Asymptomatic infectious, Hospitalized, etc. (Blackwood & Childs, 2018). Such states may also include behaviours of the population segments, including those produced via interventions such as quarantining (Maier & Brockmann, 2020) and wearing masks (Ngonghala et al., 2020). The SIR model has also been extended to incorporate the age structure in the contact matrices (Walker et al., 2020). Compartmental models are popular because they can be designed to frame the essential parts of a question and to work with reduced amounts of data for calibration. By varying parameters such as time between cases, average rate an individual can infect another, and the time infected individual can recover, researchers can estimate the case increase, as well as other properties of the epidemic. For instance, during the COVID-19 epidemic, the effective reproduction number R, or average number of secondary cases per infectious case in a population made up of both susceptible and non-susceptible hosts, has been closely watched and estimated in different affected countries, providing an important characterization of the disease's spread (D'Arienzo &

² https://sdgs.un.org/goals/goal3

Coniglio, 2020). This classic model has been recently challenged and improvements have been proposed. For instance, the assumption that any individual may contact and thus infect any other in a population (*homogeneous mixing*) has been shown to be oversimplification of the way people interact in reality; instead, considering other information, such as differential susceptibility by age, may improve the models models (Q.-H. Liu et al., 2018).

Further, the availability of large-scale data has allowed scholars to model the realworld networks more accurately. The effect of network structure has been studied in the context of epidemic spreading velocity (Cui et al., 2014) and size (Y. Liu et al., 2016; Wu et al., 2015) and thresholds (Silva et al., 2019). Pandemic outbreaks have been found to be supported in networks with high assortativity (Moreno et al., 2003) and those having community structures (Z. Liu & Hu, 2005). The plethora of data has also allowed the application of agent-based models (ABMs) which attempt to capture empirical socio-demographic characteristics such as household's sizes and compositions, however at a larger computational cost. Such models have been used to incorporate empirical knowledge about contact rates within and between age groups (Ogden et al., 2020) and comorbidities (Wilder et al., 2020). Most such models are built using known population statistics, such as the ABM built to simulate disease evolution in France in order to evaluate the effectiveness of COVID-19 lockdowns, physical distancing, and mask-wearing (Hoertel et al., 2020). Alternatively, contact tracing data has been used to build detailed community network approximations, such as one built for Boston, by considering anonymized GDPR-compliant mobile location data in combination with 83,000 places from Foursquare (Aleta et al., 2020). To make sure data sparsity does not result in individual privacy violations, the authors use a probabilistic approach to measure co-presence. Thus, ABMs have been useful in furthering our understanding of the changes to contact networks and their impact on disease transmission.

Fine-grained mobile phone data has been used to estimate population movements affecting the spread of influenza-like illness (ILI) predating COVID-19. In Tizzoni et al. (2014), the data comes as a set of phone calls georeferenced to the cellphone tower. The authors estimate that a user's most frequent location in the data is their residence and second-most frequent is the place of employment. Usually obtained via extensive (and expensive) surveys, such information is revolutionizing disease modelling on both local and global scales. Beyond phone records, internet data has also been used to monitor mobility. These works show the possibility for large corporations to surface anonymized, aggregated, and differentially private data in order to assist public health researchers and decision-makers. These include Google COVID-19 Community Mobility Reports (Google, 2021a), Apple Mobility Trends Reports (Apple., 2021), and Facebook Disease Prevention Maps (Facebook, 2021b), all of which aggregate the massive amounts of information their platforms collect about the location of their users. All three resources have been used to gauge the changes in mobility of during the COVID-19 lockdowns (Mejova & Kourtellis, 2021; Shepherd et al., 2021; Woskie et al., 2021). However, if one wants to obtain a more nuanced understanding of contact networks, wearable technologies can be used to detect face-to-face interactions within, say, an organization or a building. Unobtrusive sensors have been used to detect close proximity interactions at 1.5 m in order to reveal the interaction patterns among healthcare workers and patients in a hospital (Vanhems et al., 2013), as well as at an academic conference (Smieszek et al., 2016) and within several households in Kenya (Kiti et al., 2016). Large-scale proximity sensors were later used by many governments during the COVID-19 epidemic through passive contact tracing apps, which use anonymous identifiers to remember devices which were in a close proximity of a person and which can notify their users in case somebody within their contact history has been found to be COVID-positive (Barrat et al., 2020).

But before the disease can be tracked, its very presence needs to be detected. Computational social science presents several unprecedented data sources that enable researchers to "now-cast" disease as it moves through the population. As mentioned, web search data has been used to monitor ILI symptoms (Ginsberg et al., 2009) and is still used for many others. However, one does not need to be a Google employee to perform such research, as aggregated search data is surfaced by the company via Google Search Trends (Google, 2021b), which has been used to track anything from Lyme disease (Kapitány-Fövény et al., 2019) to type 2 diabetes (Tkachenko et al., 2017). Of course, other dynamic social media have been used to track disease, including Twitter, Reddit, and Sina Weibo, all of which have been used to track non-communicable diseases as well. Beyond observation, self-reported data can be obtained from participatory surveillance systems, such as InfluenzaNet (Koppeschaar et al., 2017), which collects influenza-related information from thousands of volunteers from countries around the EU.

Both algorithmic and data advances described above come with many caveats which both the scientific and policy communities are yet to tackle effectively. As machine learning and other modelling algorithms become more complex, difficulties in communicating their benefits and—more importantly—limitations to those outside the initiated trained practitioners result in misunderstandings about the certainty of the predictions and limits of their applications, leading to a limited deployment in the field. However, the solution may not lie in a more detailed description of the algorithms, but in the clarification of their merits, such that we can be determined whether their performance warrants their integration in the decisionmaking process of policymaking. One could take a page from the social science "reproducibility crisis" (Camerer et al., 2018) which illustrated the bias toward significant, positive, and theoretically neat results at the cost of valid, generalizable insights. Several actions, including the Social Sciences Replication Project (SSRP), the Reproducibility Project: Psychology (RPP), and the Experimental Economics Replication Project (EERP), have been organized to provide increased rigor to the insights on important theories and results in each field. Beyond reproducibility, integration of new methodologies should be tested in prediction competitions, such as CDC's FluSight, a competition that brings together researchers and industry leaders to forecast the timing, peak, and intensity of the flu season (Centers for Disease Control and Prevention., 2021). Another ongoing effort is the ECDC's European Covid-19 Forecast Hub which collates and combines short-term forecasts of COVID-19 generated by different independent modelling teams across Europe and makes available a near-term future trajectory of the pandemic (European Centre for Disease Control and Prevention (ECDC), 2021). The legitimacy afforded by such efforts would encourage the data owners (e.g. internet/technology companies including social media websites and phone companies) to contribute datasets that would level the playing field between well-funded and smaller players. It is especially important to solicit both algorithmic and expert (human) predictions in order to provide a baseline for comparison, as it has been shown that people tend to distrust algorithms faster when they make mistakes, compared to when humans do the same (Dietvorst et al., 2015). Increased transparency in the way epidemiological studies are designed, the kind of data they use, and—crucially—their predictions ahead of the target date are all likely not only to clarify the potential impact of the new methods on public health but also to unify the field under a set of common goals (Miguel et al., 2014).

This proposal will hopefully address several other critiques. Legitimizing and clearly describing the uses of data would give a greater transparency to the secondary use of data, greater oversight over anonymization standards, and aggregate statistics of its biases. Biases in data collection have been a constant critique of scientific endeavours; however, it may be even easier to gloss over biases in big datasets, but it has been shown that even large datasets of internet or technology users have substantial biases in terms of demographics, wealth, and technological access (Hargittai, 2020; Yom-Tov, 2019). Sampling biases limit the generalizability of the scientific studies. As such biases tend to underrepresent those coming from more disadvantaged backgrounds and locales, systematic testing of the algorithms on different populations would provide a quantifiable measure of the change in performance across groups of interest (Olteanu et al., 2019). The peculiarities of the digital platforms provide another constraint, including the affordances provided by each website, as well as the peculiar user base and culture. For instance, the privacy and identification limitations on Facebook distinguish it from more open platforms, like Twitter, or community-oriented ones, like Reddit, resulting in differences of information disclosure and propagation. The very timing of the studies imposes biases specific to the time period selected for the analysis (for instance, 2020 will likely be a special year in many datasets), making some observations unique to the contemporary societal, technological, and public health situation. To address some of these problems, scientists must be encouraged to publish replication studies, as well as to extend them into long-term projects, in order to test the models initially proposed on different data and time spans. Further, establishing data partnerships addressing important public health concerns will insure the infrastructure is in place in case a crisis, such as the COVID-19 epidemic, strikes.

15.3.2 Non-communicable Diseases

As medicine advanced against infectious diseases, non-communicable diseases have become the leading causes of death and illness throughout developed and developing

world. Many of such conditions, including obesity and the overweight, diabetes, and cardiovascular complications, have a strong "lifestyle" component, wherein the daily activities of the population accumulate to contribute to worsening outcomes. CSS provides a unique view of such behaviours, using the digital traces left through these daily activities such as social media posts, business check-ins, web searches, use of applications, and many others. Behaviours around food consumption and nutrition have been studied using Twitter (Abbar et al., 2015), Instagram (Mejova et al., 2015), as well as large datasets of grocery purchases (Aiello et al., 2019). Often, natural language processing (NLP) tools are used to process the text obtained from many internet users or deep machine learning (ML) models to "recognize" relevant objects in the shared images in order to understand the daily behaviours of the internet users. Crucially, these activities can be put into a cultural context to better understand the societal, economic, and psychological forces shaping these daily decisions, much as proposed by Weiss as "cultural epidemiology" (Weiss, 2001) that combines quantitative and qualitative methodologies. For instance, large datasets of recipes have been examined in order to establish a network of flavours and ingredients across countries and relate it to the health outcomes of different locales (Sajadmanesh et al., 2017). The relationship between economic deprivation on diet in the USA has shown that those living in "food deserts" mention food that is higher in fat, cholesterol, and sugar than otherwise (De Choudhury, Sharma, et al., 2016). Further, specialized apps and wearables are used to monitor physical activity. For example, a study of running tracking app data (Aral & Nicolaides, 2017) aimed to understand the role of social interaction and comparison on the duration of one's run. However, some researchers aim to go beyond behavioural profiling and use internet search data to detect those potentially having serious illness. A team used search query logs to first identify users who mentioned having a diabetes diagnosis and compare them to a control group (Hochberg et al., 2019). Researchers were able to predict whether a user will be searching for diabetes-related words from their previous queries with a positive predictive value of 56% at a false-positive rate of 1% at up to 240 days before they mention the diagnosis. In general, it was found that people tend to search about symptoms some time before they are diagnosed with the underlying condition (Hochberg et al., 2020), especially if the symptoms are serious. Yet more data is available to monitor disease on a population level via information surfaced by the advertising systems of large social media platforms. For instance, Facebook allows potential advertisers to run detailed queries on their target audience, specifying their demographics, precise location, language, and interests (which span health concerns, activities, hobbies, worldviews, and many more categories) (Facebook, 2021a). These can then be used as a kind of "digital census" to quantify awareness of health-related topics and behaviours related to noncommunicable diseases within well-defined demographic groups across fine and broad geographies (Mejova, Weber, et al., 2018). Compared to traditional surveybased monitoring, the above studies provide unobtrusive, real-time, and extremely rich sources of behavioural observation. Especially on social media, the users are self-motivated to share their meals and activities, to annotate them with geographic and other metadata, and to interact with other posts. Although suffering from social

desirability bias, in combination with other consumption statistics, social media and app use data provide important signals about the social and psychological context of health-related behaviours.

Further, non-communicable disease interventions can be studied on a personal level while delivered through a myriad of technologies. Integration of smartphones with user-generated content is leading to sophisticated personalized interventions aiming at motivating the users to increase their physical activity level (Harrington et al., 2018; op den Akker et al., 2014). Different messaging strategies have been explored including personalized exercise recommendation (Tseng et al., 2015), also employing machine learning via supervised learning (Hales et al., 2016; Marsaux et al., 2016) and reinforcement learning (Rabbi et al., 2015; Yom-Tov et al., 2017). Others help users find exercise partners (Hales et al., 2016) and provide educational materials (Short et al., 2017) and emotional support (Vandelanotte et al., 2015). The applications have been embraced by the governments and businesses worldwide. For instance, UK's National Health Service promotes an Active 10 app that encourages everyone to have a brisk walk and for those ready for a bigger challenge has *Couch* to 5K app for beginner runners (National Health Service., 2021). India's Ministry of Youth Affairs and Sports launched its *Fit India* app to help its populous keep track of their fitness goals, water intake, and sleep (Play Store., 2021). Social media is, of course, another popular outlet for public health outreach. Many associations, such as the National Eating Disorders Association in the USA, run annual health awareness campaigns on different social media channels, making it possible to measure the impact of their campaigns on the sustainability of the attention to the topic and other subsequent behaviours expressed by their audience (Mejova & Suarez-Lledó, 2020). To assist in the efforts, some researchers focus on which influencers and content (especially contagious "memes") are particularly successful in attracting an audience (Kostygina et al., 2020) or how to better identify the relevant users to target (Chu et al., 2019).

Although the above studies provide a valuable context to the ongoing epidemics of non-communicable diseases, and potential avenues to communicate about them, mostly observational studies usually fail to reach the threshold for causal insight. Often large datasets lack the information on important confounders that may affect the outcome of the study. For instance, while comparing healthrelated interests expressed by Facebook users to rates of obesity, diabetes, and alcoholism, researchers have found that unrelated (or "placebo") interests, such as those in entertainment or technology, also had substantial correlation with the rates of disease (Mejova, Weber, et al., 2018). Some attempt to improve the quality of their models by employing instrumental variables, especially when the explanatory variable of interest is correlated with the error term. Weather is a popular instrumental variable, as it is often not related to the dependent variable, but may have some relationship with the independent ones. In their study of social contagion in a community of runners, the authors used the weather at one person's location as an instrumental variable when modelling the running behaviour of another (Aral & Nicolaides, 2017). They show that without the corrections, the effect would have been overestimated by 71-82%. The inability to acquire multidimensional data that has important confounders (which are often demographics, protected by numerous privacy regulations) has an additional effect of hiding the unequal relevance of the ongoing work to those less represented in these datasets. Inferring sensitive information, including age, gender, and location, may be possible from some sources of data, but such activity may both break the privacy of the platform and violate the protections imposed by the EU General Data Protection Regulation (GDPR). It is thus imperative to engage legitimate stakeholders who will negotiate controlled releases of highly detailed data for research on pressing topics and especially provide input during policy changes when a "natural experiment" may take place. Policymakers may also want to explicitly outline the under-served populations they would like to focus on, thus encouraging the creation of datasets around groups that are not yet captured in currently available data. For instance, India's efforts in the National Mission for Empowerment of Women (NMEW) may be augmented by encouraging the monitoring of technology use through available data (Mejova, Gandhi et al., 2018). Alternatively, access to care can be monitored using online tools, such as those for women's health services (Dodge et al., 2018) across the USA.

15.3.3 Mental Illness and Suicide

An especially vulnerable population that has been extensively studied by CSS in the context of epidemiology is people with diagnosed mental illness, or those simply expressing mental distress, alongside those who vocally contemplate suicide. The anonymity and social support provided by the internet forums and websites allow many to express feelings and thoughts which may be difficult to evoke using standard public health methods like surveys and medical records. The pervasive use of social media, including on mobile devices, allows users to post instantly during the moments of mental distress and for some to integrate digital platforms into their coping mechanisms. Communities around eating disorders (anorexia, bulimia, etc.) (Stewart et al., 2017; Yom-Tov et al., 2012), depression (De Choudhury et al., 2013; Reece & Danforth, 2017), and drug abuse (Kazemi et al., 2017) and recovery (Chancellor et al., 2019) are providing valuable insights in the way people experience these conditions, seek and provide support, and even provide practical advice. For instance, by combining automated machine learning classification and text processing techniques with clinical expertise, researchers have used the Reddit opioid addiction recovery forums to discover alternative treatments that the users share and discuss (Chancellor et al., 2019). It is also possible to monitor the progression of mental illness to serious suicide ideation by examining suicide prevention forums (De Choudhury, Kiciman et al., 2016), as well as studying web search patterns (Adler et al., 2019). Studies of search engine usage have been able to confirm behavioural signs of people with autism, for instance, finding that users who have self-stated that they have autism spend less time examining image results (Yechiam, Yom-Tov et al., 2021). Whereas most studies rely on self-declaration of diagnosis, some studies use social media to better understand those who have been confirmed to be clinically diagnosed. Facebook posts of patients diagnosed with a primary psychotic disorder have been analysed to find predictors of a future psychotic relapse (Birnbaum et al., 2019).

However, the very fact that self-expression of mental distress may come before official diagnosis makes such research struggle with construct validity, that is, what exactly is being measured, and how robust it is in clinical terms. Reviews of literature on mental health status on social media show that few use the definitions and theories developed in the clinical setting to define, for instance, the conditions of "anxiety" or "depression" that are being tracked (Chancellor & De Choudhury, 2020). Whether mentions of disorders on social media capture users who are struggling with them, merely interested in the topic, or even misusing the terms is an important question to answer before these methods can be applied to the clinical setting. It is imperative to foster a closer collaboration between the medical establishment and researchers attempting to contribute to the epidemiology of conditions possibly discussed in user-generated data. From the CSS research community's side, it is important to rigorously define the cohorts of interest and follow clinically validated diagnostic procedures (Ernala et al., 2019) when studying new sources of data and methods for identifying those potentially struggling with mental illness. However, it is also desirable to have the medical community to acknowledge these new sources of information as an additional signal that should be clinically studied and which may play a role in official diagnostic (and possibly treatment) frameworks. As mentioned earlier, methods based on alternative data sources may play a role in the profiling of future recruits for studies, potentially expanding their reach beyond those already in the medical system.

15.3.4 Beliefs, Information, and Misinformation

User-generated data provides yet another unique context around health and disease: the dynamics of individual's knowledge, opinion, and belief and their interactions with various information sources that shape these important precursors to behaviour. The quality of medical information available to people on social media and through web search can be evaluated using big data NLP tools and in collaboration with area experts. YouTube videos have been found to be some of the worst offenders in terms of advocating methods proven to be harmful or having no scientific basis (Madathil et al., 2015). Twitter (Rosenberg et al., 2020), Reddit (Jang et al., 2019), and Pinterest (Guidry & Messner, 2017) all have been examined for links to potentially harmful health advice. One of the most serious problems is the anti-vaccination movement that has been strengthening in both developed and developing countries. Twitter data has proven to be useful in explaining some variation in the vaccine coverage rates, as reported by the immunization monitoring system of WHO (Bello-Orgaz et al., 2017). Classifying whether social media users support or oppose vaccinations has been shown to be feasible, both using deep learning on the

posted text and images (Wang et al., 2020) and using network algorithms on the conversation network (Cossard et al., 2020). However, it is in the more specialized websites, such as the discussion forums for parents, that give space to those who are hesitant and are in the process of making healthcare decisions for themselves or their family. There, researchers can find lists of concerns, previous experiences, and information seeking, as well as testimonials about the experiences with the medical establishment (Betti et al., 2021).

Further, internet captures myriad interactions with medical services and consequences of health interventions. Social media has been used extensively for *pharmacovigilance*, discovering drug side effects (Alvaro et al., 2015), drug interactions (Correia et al., 2016), and recreational drug use (Deluca et al., 2012) and even uncovering illicit online pharmacies (Katsuki et al., 2015). Patient experiences can be found on business review websites (Rastegar-Mojarad et al., 2015), as well as general-purpose social media, where communities can discuss their perceptions of treatment (Booth et al., 2019; Hswen et al., 2020). Super-utilizers of healthcare services have also been studied on social media in order to inform online social support interventions and complement offline community care services (Guntuku et al., 2021), and efforts have been made to integrate patient experiences in online discussions into customer satisfaction and service quality measures (Albarrak & Li, 2018).

As more and more people use internet and social media as a source of medical knowledge and advice, as well as social support, understanding how this information is translated into behaviours and life choices is an increasingly urgent research direction. Although the detection of cyberbullying and other negative speech on social media is an active research direction (Chatzakou et al., 2019), ethical concerns prevent the integration of user profiling and targeting in mental health interventions. However, health misinformation has been acknowledged to be a parallel pandemic in the COVID-19 era, and concerted efforts are ongoing in monitoring and tackling potentially harmful information (World Health Organization, 2021a). In this sphere, CSS will continue to play an important role by providing the tools for the analysis of new social and information sharing platforms that are increasingly permeating the information landscape.

15.4 The Way Forward

Epidemiology was one of the first of the sciences to use large datasets, and thus, it is in a natural position to take advantage of the latest developments in digitization, big data, and computing methods. The year 2020 has forced the field to mobilize its best resources to address the COVID-19 pandemic and put in stark light the challenges facing the field. The silver lining of this dark cloud could be an understanding of the necessary steps in bringing digital epidemiology into the policy sphere, making it agile and relevant in a fast-moving globalized world. The COVID-19 pandemic has imparted an important component to the epidemiological field—a clarity of vision. It has shown in a stark contrast the cost of indecision and the global repercussions on the lives and economies and forced the realities of a global pandemic to the public and the governmental attention. It has also revealed weaknesses in the current health policy structures, the slow response of the governments to the WHO's messaging, and disarray in the case tracking and reporting standards. Already, actions are in place to remedy these weaknesses. Attempts are being made to formalize the government responses through treaties and international agreements (though enforcing such agreements remains a struggle) (Maxmen, 2021). Partnerships are being forged, and large companies released detailed datasets of user activity and mobility to aid in monitoring and modelling (National Institutes of Health., 2021).

Such clarity of vision is necessary to improve the impact of digital epidemiology also in other spheres. The UN Sustainable Development Goals (SDGs)³ provide a general prioritization for the health and well-being challenges, but these must be defined clearly in order to encourage the building of tools and partnerships. One such effort is the European Data Space, which aims to legitimize and operationalize the data usage across the member states while complying with its established privacy regulations.⁴ Another is the WHO Hub for Pandemic and Epidemic Intelligence which aims to build a "global trust architecture" that will encourage greater sharing of data through addressing numerous aspects: "governance, legal frameworks and data-sharing agreements; data solidarity, fairness and benefits sharing; transparency about how pandemic and epidemic intelligence outputs are used; openness of technology solutions and artificial intelligence applications; security of data; combating misinformation and addressing infodemics; privacy by design principles; and public participation and people's data literacy" (World Health Organization, 2021b,c). Additionally, the One Health movement, supported by the WHO, emphasizes the collaboration between disparate domains to accomplish a systems-level perspective on problems such as antibiotic resistance (World Health Organization., 2017). These ambitious projects are a response to a complex problem that involves many parties, some of which only recently began weighing the benefits and dangers of massive surveillance for the greater good.

Several important steps need to be taken in order to engage all major parties involved. First, civil society must be educated in the basics of digital literacy, data privacy, and its governance in order to ensure the users of technologies contributing to the big data revolution provide truly informed consent. For instance, the EU has proposed the Digital Competence Framework that comprises not only information literacy but also skills in communication, digital content creation, safety, and problem-solving (EU Science Hub., 2021). Second, the professionals coming from different civic, academic, and policy silos must be brought together and upskilled to legibly communicate about the role of data in public health. For instance, efforts

³ https://sdgs.un.org/goals

⁴ https://ec.europa.eu/health/ehealth/dataspace_en

such as the Lagrange Fellowships in Italy (Fondazione CRT., 2019); the Data Science for Social Good Fellowship in Chicago, USA (University of Chicago., 2021); and the Data Fellowship at the OCHA Centre for Humanitarian Data (Centre for Humanitarian Data, 2021) are excellent efforts to impart data science skills in the next generation of humanitarians, epidemiologists, and academics. Institutionally, the normalization of building teams that incorporate data literacy (and analytics skills, if possible) is an ongoing process that is only recently being supported by educational resources. Third, the governance of technology giants that own much of the data necessary for monitoring and modelling disease must be kept clear and up-to-date considering the latest technological developments. Interestingly, during the COVID-19 efforts to build contact tracing apps, it was the corporations (Apple and Google) that refused to implement features that would threaten the privacy of their users (privacy being an important feature of their services) (Meyer, 2021). However, one must not rely on the businesses to maintain ethical standards of data use, which must be carefully negotiated before the next disaster strikes.

Much of this chapter describes the impressive accomplishments by the academic researchers in the fields of disease monitoring, modelling, prediction, and contextualization. However, to bring these tools to the policymakers' table, they must be robust, vetted, and available on demand. Additional organization is necessary to establish a well-defined set of problems for the community to tackle and to provide legitimacy in order to foster data exchange to support research. Standardizing the tasks (such as flu season prediction), metrics, available data, and benchmarks will allow for an increased accountability and reproducibility of academic endeavours that go beyond publication peer review. Such tasks should be defined in collaboration with the policymakers in order to align the priorities with the societal needs and system outputs with the information needs. The way Netflix Prize has invigorated the recommender systems community (Netflix, 2009) and Google Flu Trends spurred interest in the digital disease tracking (Google., 2014), ambitious competitions not only would provide clarity of vision for the field but would also be able to direct the research agenda to under-served areas and communities. It would be beneficial if the collaborative efforts described above would include a space for the academics and researchers to tackle specific problems within an evaluation framework that produces benchmark datasets and reproducible methods, beyond scientific publications.

Finally, the technological development will continue revolutionizing the field, spurring debate on additional policy considerations. The advances in deep machine learning are allowing to process speech, images, and video at scale and are already being used for plant (Ferentinos, 2018) and human (Li et al., 2020) disease detection. The rise of confidential computing, wherein user data is isolated and protected on the user's device, and only trusted operations can be run on it, eliminates the need to transfer the data for processing elsewhere (Rashid, 2020). The negotiation between the new potential insights and the cost to the society will require thoughtful, informed, and urgent consideration.

References

- Abbar, S., Mejova, Y., & Weber, I. (2015). You tweet what you eat: Studying food consumption through twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3197–3206).
- Adler, N., Cattuto, C., Kalimeri, K., Paolotti, D., Tizzoni, M., Verhulst, S., Yom-Tov, E., & Young, A. (2019). How search engine data enhance the understanding of determinants of suicide in india and inform prevention: Observational study. *Journal of medical internet research*, 21(1), e10179.
- Aiello, L. M., Schifanella, R., Quercia, D., & Del Prete, L. (2019). Large-scale and high-resolution analysis of food purchases and health outcomes. *EPJ Data Science*, 8(1), 1–22.
- Albarrak, A., & Li, Y. (2018). Quality and customer satisfaction health accessibility framework using social media platform. In *Proceedings of the 51st Hawaii International Conference on System Sciences*.
- Aldridge, J., & Décary-Hétu, D. (2016). Hidden wholesale: The drug diffusing capacity of online drug cryptomarkets. *International Journal of Drug Policy*, 35, 7–15.
- Aleta, A., Martin-Corral, D., Pastore y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N. E., Halloran, M. E., Longini Jr, I. M., Merler, S., et al. (2020). Modelling the impact of testing, contact tracing and household quarantine on second waves of covid-19. *Nature Human Behaviour*, 4(9), 964–971.
- Althoff, T., White, R. W., & Horvitz, E. (2016). Influence of pokémon go on physical activity: Study and implications. *Journal of Medical Internet Research*, 18(12), e315.
- Alvaro, N., Conway, M., Doan, S., Lofi, C., Overington, J., & Collier, N. (2015). Crowdsourcing twitter annotations to identify first-hand experiences of prescription drug use. *Journal of Biomedical Informatics*, 58, 280–287.
- Apple. (2021). Mobility trends reports. Accessed 1 Sep 2021.
- Aral, S., & Nicolaides, C. (2017). Exercise contagion in a global social network. *Nature Communications*, 8(1), 1–8.
- Azad, S., & Devi, S. (2020). Tracking the spread of covid-19 in india via social networks in the early phase of the pandemic. *Journal of Travel Medicine*, 27(8), taaa130.
- Barrat, A., Cattuto, C., Kivelä, M., Lehmann, S., & Saramäki, J. (2020). Effect of manual and digital contact tracing on covid-19 outbreaks: A study on empirical contact data. *Journal of the Royal Society Interface*, 18(178), 20201000.
- Bello-Orgaz, G., Hernandez-Castro, J., & Camacho, D. (2017). Detecting discussion communities on vaccination in twitter. *Future Generation Computer Systems*, 66, 125–136.
- Betti, L., De Francisci Morales, G., Gauvin, L., Kalimeri, K., Mejova, Y., Paolotti, D., & Starnini, M. (2021). Detecting adherence to the recommended childhood vaccination schedule from user-generated content in a us parenting forum. *PLoS Computational Biology*, 17(4), e1008919.
- Birnbaum, M. L., Ernala, S. K., Rizvi, A., Arenare, E., Van Meter, A., De Choudhury, M., & Kane, J. M. (2019). Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from facebook. *NPJ Schizophrenia*, 5(1), 1–9.
- Bjørnstad, O. N., Shea, K., Krzywinski, M., & Altman, N. (2020). Modeling infectious epidemics. *Nature Methods*, 17(5), 455–456.
- Blackwood, J. C., & Childs, L. M. (2018). An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics* 5, 195–221.
- Bonnevie, E., Rosenberg, S. D., Kummeth, C., Goldbarg, J., Wartella, E., & Smyser, J. (2020). Using social media influencers to increase knowledge and positive attitudes toward the flu vaccine. *Plos One*, 15(10), e0240828.
- Booth, A., Bell, T., Halhol, S., Pan, S., Welch, V., Merinopoulou, E., Lambrelli, D., & Cox, A. (2019). Using social media to uncover treatment experiences and decisions in patients with acute myeloid leukemia or myelodysplastic syndrome who are ineligible for intensive chemotherapy: Patient-centric qualitative data analysis. *Journal of Medical Internet Research*, 21(11), e14285.

- Budd, J., Miller, B. S., Manning, E. M., Lampos, V., Zhuang, M., Edelstein, M., Rees, G., Emery, V. C., Stevens, M. M., Keegan, N., Short, M. J., Pillay, D., Manley, E., Cox, I. J., Heymann, D., Johnson, A. M., & McKendry, R. A. (2020). Digital technologies in the public-health response to covid-19. *Nature medicine*, 26(8), 1183–1192.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., ... Wu, H. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644.
- Carlson, S., Dey, A., & Beard, F. (2020). An evaluation of the 2016 influenza vaccination in pregnancy campaign in nsw, australia. *Public Health Res Pract*, 30(1), pii–29121908.
- Centers for Disease Control and Prevention. (2021). *Flusight: Flu forecasting*. Accessed 1 Sep 2021.
- Centre for Humanitarian Data. (2021). Data fellows programme. Accessed 18 Sep 2021.
- Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. NPJ Digital Medicine, 3(1), 1–11.
- Chancellor, S., Nitzburg, G., Hu, A., Zampieri, F., & De Choudhury, M. (2019). Discovering alternative treatments for opioid use recovery using social media. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–15).
- Chatzakou, D., Leontiadis, I., Blackburn, J., Cristofaro, E. D., Stringhini, G., Vakali, A., & Kourtellis, N. (2019). Detecting cyberbullying and cyberaggression in social media. ACM Transactions on the Web (TWEB), 13(3), 1–51.
- Cheng, Q., Li, T. M., Kwok, C.-L., Zhu, T., & Yip, P. S. (2017). Assessing suicide risk and emotional distress in chinese social media: A text mining and machine learning study. *Journal* of Medical Internet Research, 19(7), e243.
- Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., y Piontti, A. P., Mu, K., Rossi, L., Sun, K., et al. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, *368*(6489), 395–400.
- Chu, K.-H., Colditz, J., Malik, M., Yates, T., & Primack, B. (2019). Identifying key target audiences for public health campaigns: Leveraging machine learning in the case of hookah tobacco smoking. *Journal of Medical Internet Research*, 21(7), e12443.
- Coppersmith, G., Harman, C.,& Dredze, M. (2014). Measuring post traumatic stress disorder in twitter. In *Eighth International AAAI Conference on Weblogs and Social Media*.
- Correia, R. B., Li, L., & Rocha, L. M. (2016). Monitoring potential drug interactions and reactions via network analysis of instagram user timelines. In *Biocomputing 2016: Proceedings of the Pacific Symposium* (pp. 492–503)
- Cossard, A., Morales, G. D. F., Kalimeri, K., Mejova, Y., Paolotti, D., & Starnini, M. (2020). Falling into the echo chamber: The italian vaccination debate on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 14, pp. 130–140).
- Cui, A.-X., Wang, W., Tang, M., Fu, Y., Liang, X., & Do, Y. (2014). Efficient allocation of heterogeneous response times in information spreading process. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(3), 033113.
- Cunha, T. O., Weber, I., Haddadi, H., & Pappa, G. L. (2016). The effect of social feedback in a reddit weight loss community. In *Proceedings of the 6th International Conference on Digital Health Conference* (pp. 99–103).
- D'Arienzo, M., & Coniglio, A. (2020). Assessment of the sars-cov-2 basic reproduction number, r0, based on the early phase of covid-19 outbreak in italy. *Biosafety and Health*, 2(2), 57–59.
- De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In *Seventh International AAAI Conference on Weblogs and Social Media*.
- De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., & Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the* 2016 CHI Conference on Human Factors in Computing Systems (pp. 2098–2110).

- De Choudhury, M., Sharma, S., & Kiciman, E. (2016). Characterizing dietary choices, nutrition, and language in food deserts via social media. In *Proceedings of the 19th Acm Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1157–1170).
- Deluca, P., Davey, Z., Corazza, O., Di Furia, L., Farre, M., Flesland, L. H., Mannonen, M., Majava, A., Peltoniemi, T., Pasinetti, M., et al. (2012). Identifying emerging trends in recreational drug use; outcomes from the psychonaut web mapping project. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 39(2), 221–226.
- Dicker, R. C., Coronado, F., Koo, D., & Parrish, R. G. (2006). Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics. Self-study course. Stephen B. Thacker CDC Library collection.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114.
- Dodge, L. E., Phillips, S. J., Neo, D. T., Nippita, S., Paul, M. E., & Hacker, M. R. (2018). Quality of information available online for abortion self-referral. *Obstetrics and Gynecology*, 132(6), 1443.
- Ernala, S. K., Birnbaum, M. L., Candan, K. A., Rizvi, A. F., Sterling, W. A., Kane, J. M., & De Choudhury, M. (2019). Methodological gaps in predicting mental health states from social media: Triangulating diagnostic signals. In *Proceedings of the 2019 Chi Conference on Human Factors in Computing Systems* (pp. 1–16).
- EU Science Hub. (2021). The digital competence framework 2.0. Accessed 19 Sep 2021.
- European Centre for Disease Control and Prevention (ECDC). (2021). European covid-19 forecast hub. Accessed 1 Sep 2021.
- European Commission. (2021). European health data space. Accessed 1 Sep 2021.
- Facebook. (2021a). Ads manager. Accessed 1 Sep 2021.
- Facebook. (2021b). Disease prevention maps. Accessed 1 Sep 2021.
- Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311–318.
- Fondazione CRT. (2019). Borse lagrange. Accessed 18 Sep 2021.

Fontana, M., & Guerzoni, M. (2023). Modeling complexity with unconventional data: Foundational issues in computational social science. In Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., Vespe, M. (Eds.), *Handbook of computational social science for policy*. Springer.

- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012–1014.
- Google. (2014). *Google flu trends*. Accessed 18 Sep 2021. Google. (2021a). *Covid-19 community mobility reports*. Accessed 1 Sep 2021.
- Google. (2021b). *Trends*. Accessed 1 Sep 2021.
- Guidry, J., & Messner, M. (2017). Health misinformation via social media: The case of vaccine safety on pinterest. In Social media and crisis communication (pp. 267–279). Routledge.
- Guntuku, S. C., Klinger, E. V., McCalpin, H. J., Ungar, L. H., Asch, D. A., & Merchant, R. M. (2021). Social media language of healthcare super-utilizers. *NPJ Digital Medicine*, 4(1), 1–6.
- Hales, S., Turner-McGrievy, G., Fahim, A., Freix, A., Wilcox, S., Davis, R. E., Huhns, M., & Valafar, H. (2016). A mixed-methods approach to the development, refinement, and pilot testing of social networks for improving healthy behaviors. *JMIR Human Factors*, 3(1), e4512.
- Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. Social Science Computer Review, 38(1), 10–24.
- Harrington, C. N., Wilcox, L., Connelly, K., Rogers, W., & Sanford, J. (2018). Designing health and fitness apps with older adults: Examining the value of experience-based co-design. In *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies* for Healthcare (pp. 15–24).
- Hochberg, I., Allon, R., & Yom-Tov, E. (2020). Assessment of the frequency of online searches for symptoms before diagnosis: Analysis of archival data. *Journal of Medical Internet Research*, 22(3), e15065.

- Hochberg, I., Daoud, D., Shehadeh, N., & Yom-Tov, E. (2019). Can internet search engine queries be used to diagnose diabetes? Analysis of archival search data. *Acta Diabetologica*, 56(10), 1149–1154.
- Hoertel, N., Blachier, M., Blanco, C., Olfson, M., Massetti, M., Rico, M. S., Limosin, F., & Leleu, H. (2020). A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature Medicine*, 26(9), 1417–1421.
- Hswen, Y., Zhang, A., Sewalk, K. C., Tuli, G., Brownstein, J. S., & Hawkins, J. B. (2020). Investigation of geographic and macrolevel variations in LGBTQ patient experiences: Longitudinal social media analysis. *Journal of Medical Internet Research.*, 22, e17087.
- Jang, S. M., Mckeever, B. W., Mckeever, R., & Kim, J. K. (2019). From social media to mainstream news: The information flow of the vaccine-autism controversy in the US, Canada, and the UK. *Health Communication*, 34(1), 110–117.
- Jeffrey, B., Walters, C. E., Ainslie, K. E., Eales, O., Ciavarella, C., Bhatia, S., Hayes, S., Baguelin, M., Boonyasiri, A., Brazeau, N. F., Cuomo-Dannenburg, G., FitzJohn, R. G., Gaythorpe, K., Green, W., Imai, N., Mellan, T. A., Mishra, S., Nouvellet, P., Juliette, H., ... Riley, S. (2020). Anonymised and aggregated crowd level mobility data from mobile phones suggests that initial compliance with COVID-19 social distancing interventions was high and geographically consistent across the UK. *Wellcome Open Research*, 5, 170.
- Johnson, N. F., Velásquez, N., Restrepo, N. J., Leahy, R., Gabriel, N., El Oud, S., Zheng, M., Manrique, P., Wuchty, S., & Lupu, Y. (2020). The online competition between pro-and antivaccination views. *Nature*, 582(7811), 230–233.
- Kapitány-Fövény, M., Ferenci, T., Sulyok, Z., Kegele, J., Richter, H., Vályi-Nagy, I., & Sulyok, M. (2019). Can google trends data improve forecasting of lyme disease incidence? *Zoonoses and Public Health*, 66(1), 101–107.
- Katsuki, T., Mackey, T. K., & Cuomo, R. (2015). Establishing a link between prescription drug abuse and illicit online pharmacies: Analysis of twitter data. *Journal of Medical Internet Research*, 17(12), e280.
- Kazemi, D. M., Borsari, B., Levine, M. J., & Dooley, B. (2017). Systematic review of surveillance by social media platforms for illicit drug use. *Journal of Public Health*, 39(4), 763–776.
- Keller, S. N., Honea, J. C., & Ollivant, R. (2021). How social media comments inform the promotion of mask-wearing and other covid-19 prevention strategies. *International Journal* of Environmental Research and Public Health, 18(11), 5624.
- Kiti, M. C., Tizzoni, M., Kinyanjui, T. M., Koech, D. C., Munywoki, P. K., Meriac, M., Cappa, L., Panisson, A., Barrat, A., Cattuto, C., et al. (2016). Quantifying social contacts in a household setting of rural kenya using wearable proximity sensors. *EPJ Data Science*, 5(1), 1–21.
- Kleppmann, M. (2017). Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. O'Reilly Media.
- Koppeschaar, C. E., Colizza, V., Guerrisi, C., Turbelin, C., Duggan, J., Edmunds, W. J., Kjelsø, C., Mexia, R., Moreno, Y., Meloni, S., Paolotti, D., Perrotta, D., van Straten, E., Franco, A. O. (2017). Influenzanet: Citizens among 10 countries collaborating to monitor influenza in Europe. *JMIR Public Health and Surveillance*, 3(3), e7429.
- Kostygina, G., Tran, H., Binns, S., Szczypka, G., Emery, S., Vallone, D., & Hair, E. (2020). Boosting health campaign reach and engagement through use of social media influencers and memes. *Social Media*+ *Society*, 6(2), 2056305120912475.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science*, 343(6176), 1203–1205.
- Li, L.-F., Wang, X., Hu, W.-J., Xiong, N. N., Du, Y.-X., & Li, B.-S. (2020). Deep learning in skin disease image recognition: A review. *IEEE Access*, 8, 208264–208280.
- Liu, Q.-H., Ajelli, M., Aleta, A., Merler, S., Moreno, Y., & Vespignani, A. (2018). Measurability of the epidemic reproduction number in data-driven contact networks. *Proceedings of the National Academy of Sciences*, 115(50), 12680–12685.
- Liu, Y., Deng, Y., Jusup, M., & Wang, Z. (2016). A biologically inspired immunization strategy for network epidemiology. *Journal of Theoretical Biology*, 400, 92–102.

- Liu, Z., & Hu, B. (2005). Epidemic spreading in community networks. *EPL (Europhysics Letters)*, 72(2), 315.
- Łuksza, M., & Lässig, M. (2014). A predictive fitness model for influenza. Nature, 507(7490), 57–61.
- Madathil, K. C., Rivera-Rodriguez, A. J., Greenstein, J. S., & Gramopadhye, A. K. (2015). Healthcare information on youtube: A systematic review. *Health Informatics Journal*, 21(3), 173–194.
- Maier, B. F., & Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed covid-19 cases in China. *Science*, 368(6492), 742–746.
- Marsaux, C. F., Celis-Morales, C., Livingstone, K. M., Fallaize, R., Kolossa, S., Hallmann, J., San-Cristobal, R., Navas-Carretero, S., O'Donovan, C. B., Woolhead, C., Forster, H., Moschonis, G., Lambrinou, C.-P., Surwillo, A., Godlewska, M., Hoonhout, J., Goris, A., Macready, A. L., Walsh, M. C., ... Saris, W. H. M. (2016). Changes in physical activity following a geneticbased internet-delivered personalized intervention: Randomized controlled trial (food4me). *Journal of medical Internet research*, 18(2), e30.
- Maxmen, A. (2021). Why did the world's pandemic warning system fail when covid hit? *Nature*, 589, 499–500.
- Mejova, Y., Gandhi, H. R., Rafaliya, T. J., Sitapara, M. R., Kashyap, R., & Weber, I. (2018). Measuring subnational digital gender inequality in India through gender gaps in facebook use. In Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (pp. 1–5).
- Mejova, Y., Haddadi, H., Noulas, A., & Weber, I. (2015). # Foodporn: Obesity patterns in culinary interactions. In *Proceedings of the 5th International Conference on Digital Health 2015* (pp. 51–58).
- Mejova, Y., & Kalimeri, K. (2019). Effect of values and technology use on exercise: Implications for personalized behavior change interventions. In *Proceedings of the 27th ACM Conference* on User Modeling, Adaptation and Personalization (pp. 36–45).
- Mejova, Y., & Kourtellis, N. (2021). Youtubing at home: Media sharing behavior change as proxy for mobility around covid-19 lockdowns. In 13th Acm Web Science Conference 2021 (pp. 272– 281). Association for Computing Machinery. https://doi.org/10.1145/3447535.3462494
- Mejova, Y., & Suarez-Lledó, V. (2020). Impact of online health awareness campaign: Case of national eating disorders association. In *International Conference on Social Informatics* (pp. 192–205).
- Mejova, Y., Weber, I., & Fernandez-Luque, L. (2018). Online health monitoring using facebook advertisement audience estimates in the united states: Evaluation study. *JMIR Public Health* and Surveillance, 4(1), e7217.
- Meyer, D. (2021). Apple and google flex privacy muscles with blockage of english covid contacttracing app update. Accessed 18 Sep 2021.
- Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., Imbens, G., et al. (2014). Promoting transparency in social science research. *Science*, 343(6166), 30–31.
- Moreno, Y., Gómez, J. B., & Pacheco, A. F. (2003). Epidemic incidence in correlated complex networks. *Physical Review E*, 68(3), 035103.
- National Health Service. (2021). Better health. Accessed 1 Sep 2021.
- National Institutes of Health. (2021). Open-access data and computational resources to address covid-19. Accessed 1 Sep 2021.
- Netflix. (2009). Netflix prize. Accessed 18 Sep 2021.
- Ngonghala, C. N., Iboi, E., Eikenberry, S., Scotch, M., MacIntyre, C. R., Bonds, M. H., & Gumel, A. B. (2020). Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel coronavirus. *Mathematical biosciences*, 325, 108364.
- Obi, C. G., Ezaka, E. I., Nwankwo, J. I., & Onuigbo, I. I. (2020). Role of the epidemiologist in the containment of COVID-19 pandemic. AIJR Preprints. https://doi.org/10.21467/preprints.183
- Ogden, N. H., Fazil, A., Arino, J., Berthiaume, P., Fisman, D. N., Greer, A. L., Ludwig, A., Ng, V., Tuite, A. R., Turgeon, P., Waddell, L. A., & Wu, J. (2020). Artificial intelligence in public

health: Modelling scenarios of the epidemic of COVID-19 in Canada. *Canada Communicable Disease Report*, 46(8), 198.

- Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., De Nadai, M., Letouzé, E., Salah, A. A., Benjamins, R., Cattuto, C., Colizza, V., de Cordes, N., Fraiberger, S. P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P. N., Pivetta, F., ... Vinck, P. (2020). Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advance*, 6(23), eabc0764.
- Olteanu, A., Castillo, C., Diaz, F., & Kýcýman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, *2*, 13.
- op den Akker, H., Jones, V. M., & Hermens, H. J. (2014). Tailoring real-time physical activity coaching systems: A literature survey and model. User Modeling and User-Adapted Interaction, 24(5), 351–392. https://doi.org/10.1007/s11257-014-9146-y
- Play Store. (2021). Fit India, Accessed 1 Sep 2021.
- Rabbi, M., Aung, M. H., Zhang, M., & Choudhury, T. (2015). Mybehavior: Automatic personalized health feedback from user behaviors and preferences using smartphones. In *Proceedings of the* 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 707– 718).
- Rashid, F. Y. (2020). The rise of confidential computing: Big tech companies are adopting a new security model to protect data while it's in use-[news]. *IEEE Spectrum*, 57(6), 8–9.
- Rastegar-Mojarad, M., Ye, Z., Wall, D., Murali, N., & Lin, S. (2015). Collecting and analyzing patient experiences of health care from social media. *JMIR Research Protocols*, 4(3), e3433.
- Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6, 1–12.
- Rich, E., & Miah, A. (2017). Mobile, wearable and ingestible health technologies: Towards a critical research agenda. *Health Sociology Review*, 26(1), 84–97.
- Rosenberg, H., Syed, S., & Rezaie, S. (2020). The twitter pandemic: The critical role of twitter in the dissemination of medical information and misinformation during the covid-19 pandemic. *Canadian Journal of Emergency Medicine*, 22(4), 418–421.
- Rosenblum, S., & Yom-Tov, E. (2017). Seeking web-based information about attention deficit hyperactivity disorder: Where, what, and when. *Journal of Medical Internet Research*, 19(4), e6579.
- Sajadmanesh, S., Jafarzadeh, S., Ossia, S. A., Rabiee, H. R., Haddadi, H., Mejova, Y., Musolesi, M., Cristofaro, E. D., & Stringhini, G. (2017). Kissing cuisines: Exploring worldwide culinary habits on the web. In *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 1013–1021).
- Shepherd, H. E., Atherden, F. S., Chan, H. M. T., Loveridge, A., & Tatem, A. J. (2021). Domestic and international mobility trends in the united kingdom during the covid-19 pandemic: An analysis of facebook data. *International Journal of Health Geographics 20*, 46 (2021). https:// doi.org/10.1186/s12942-021-00299-5
- Short, C., Rebar, A., James, E., Duncan, M., Courneya, K., Plotnikoff, R., Crutzen, R., & Vandelanotte, C. (2017). How do different delivery schedules of tailored web-based physical activity advice for breast cancer survivors influence intervention use and efficacy? *Journal of Cancer Survivorship*, 11(1), 80–91.
- Silva, D. H., Ferreira, S. C., Cota, W., Pastor-Satorras, R., & Castellano, C. (2019). Spectral properties and the accuracy of mean-field approaches for epidemics on correlated powerlawnetworks. *Physical Review Research*, 1(3), 033024.
- Smieszek, T., Castell, S., Barrat, A., Cattuto, C., White, P. J., & Krause, G. (2016). Contact diaries versus wearable proximity sensors in measuring contact patterns at a conference: Method comparison and participants' attitudes. *BMC Infectious Diseases*, 16(1), 1–14.
- Stewart, I., Chancellor, S., De Choudhury, M., & Eisenstein, J. (2017). # Anorexia, # anarexia, # anarexyia: Characterizing online community practices with orthographic variation. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 4353–4361).

- Taylor, L. (2023). Data justice, computational social science and policy. In Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., Vespe, M. (Eds.), *Handbook of computational social science for policy*. Springer.
- Tizzoni, M., Bajardi, P., Decuyper, A., Kon Kam King, G., Schneider, C. M., Blondel, V., Smoreda, Z., González, M. C., & Colizza, V. (2014). On the use of human mobility proxies for modeling epidemics. *PLoS Computational Biology*, 10(7), e1003716.
- Tkachenko, N., Chotvijit, S., Gupta, N., Bradley, E., Gilks, C., Guo, W., Crosby, H., Shore, E., Thiarai, M., Procter, R., et al. (2017). Google trends can improve surveillance of type 2 diabetes. *Scientific Reports*, 7(1), 1–10.
- Tseng, J. C., Lin, B.-H., Lin, Y.-F., Tseng, V. S., Day, M.-L., Wang, S.-C., Lo, K.-R., Yang, Y.-C. (2015). An interactive healthcare system with personalized diet and exercise guideline recommendation. In 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI) (pp. 525–532).
- University of Chicago. (2021). Data science for social good summer fellowship. Accessed 18 Sep 2021.
- Van den Broeck, W., Gioannini, C., Gonçalves, B., Quaggiotto, M., Colizza, V., & Vespignani, A. (2011). The gleamviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC Infectious Diseases*, 11(1), 1–14.
- Vandelanotte, C., Short, C., Plotnikoff, R. C., Hooker, C., Canoy, D., Rebar, A., Alley, S., Schoeppe, S., Mummery, W. K., & Duncan, M. J. (2015). Tayloractive–examining the effectiveness of web-based personally-tailored videos to increase physical activity: A randomised controlled trial protocol. *BMC Public Health*, 15(1), 1020.
- Vanhems, P., Barrat, A., Cattuto, C., Pinton, J.-F., Khanafer, N., Régis, C., Kim, B.-a., Comte, B., & Voirin, N. (2013). Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PloS One*, 8(9), e73970.
- Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical challenges of big data in public health. *PLOS Computational Biology*, 11, e1003904.
- Venkatramanan, S., Sadilek, A., Fadikar, A., Barrett, C. L., Biggerstaff, M., Chen, J., Dotiwalla, X., Eastham, P., Gipson, B., Higdon, D., Kucuktunc, O., Lieber, A., Lewis, B. L., Reynolds, Z., Vullikanti, A. K., Wang, L., & Marathe, M. (2021). Forecasting influenza activity using machine-learned mobility map. *Nature communications*, 12(1), 1–12.
- Vespe, M., Iacus, S. M., Santamaria, C., Sermi, F., & Spyratos, S. (2021). On the use of data from multiple mobile network operators in europe to fight COVID-19. *Data & Policy*, 3, E9.
- Volz, E. M., Koelle, K., & Bedford, T. (2013). Viral phylodynamics. *PLoS Computational Biology*, 9(3), e1002947.
- Walker, P. G., Whittaker, C., Watson, O. J., Baguelin, M., Winskill, P., Hamlet, A., Djafaara, B. A., Cucunubá, Z., Olivera Mesa, D., Green, W., Thompson, H., Nayagam, S., Ainslie, K. E. C., Bhatia, S., Bhatt, S., Boonyasiri, A., Boyd, O., Brazeau, N. F., Cattarino, L., ... Ghani, A. C. (2020). The impact of COVID-19 and strategies for mitigation and suppression in low-and middle-income countries. *Science*, 369(6502), 413–422.
- Wang, Z., Yin, Z., & Argyris, Y. A. (2020). Detecting medical misinformation on social media using multimodal deep learning. *IEEE Journal of Biomedical and Health Informatics*, 25(6), 2193–2203.
- Weber, I., & Achananuparp, P. (2016). Insights from machine-learned diet success prediction. In Biocomputing 2016: Proceedings of the Pacific Symposium (pp. 540–551).
- Weidman, S. (2019). *Deep learning from scratch: Building with python from first principles*. O;Reilly Media.
- Weiss, M. G. (2001). Cultural epidemiology: An introduction and overview. Anthropology & Medicine, 8(1), 5–29.
- Weitz, J. S., Park, S.W., Eksin, C., & Dushoff, J. (2020). Awareness-driven behavior changes can shift the shape of epidemics away from peaks and toward plateaus, shoulders, and oscillations. *Proceedings of the National Academy of Sciences*, 117(51), 32764–32771.
- Wilder, B., Charpignon, M., Killian, J. A., Ou, H.-C., Mate, A., Jabbari, S., Perrault, A., Desai, A.N., Tambe, M., & Majumder, M. S. (2020). Modeling between population variation in

COVID-19 dynamics in Hubei, Lombardy, and New York City. Proceedings of the National Academy of Sciences, 117(41), 25904–25910.

- World Health Organization. (2017). One health. Accessed 21 Sep 2021.
- World Health Organization. (2021a). Fighting misinformation in the time of COVID-19, one click at a time. Accessed 21 Sep 2021.
- World Health Organization. (2021b). *Who hub for pandemic and epidemic intelligence*. Accessed 21 Sep 2021.
- World Health Organization. (2021c). Who, germany open hub for pandemic and epidemic intelligence in Berlin. Accessed 21 Sep 2021.
- Woskie, L. R., Hennessy, J., Espinosa, V., Tsai, T. C., Vispute, S., Jacobson, B. H., Cattuto, C., Gauvin, L., Tizzoni, M., Fabrikant, A., Gadepalli, K., Boulanger, A., Pearce, A., Kamath, C., Schlosberg, A., Stanton, C., Bavadekar, S., Abueg, M., Hogue, M., ..., Gabrilovich, E. (2021). Early social distancing policies in Europe, changes in mobility & COVID-19 case trajectories: Insights from spring 2020. *Plos one*, *16*(6), e0253071.
- Wu, Q., Fu, X., Jin, Z., & Small, M. (2015). Influence of dynamic immunization on epidemic spreading in networks. *Physica A: Statistical Mechanics and its Applications*, 419, 566–574.
- Yamada, Y., Ćepulić, D.-B., Coll-Martin, T., Debove, S., Gautreau, G., Han, H., Rasmussen, J., Tran, T. P., Travaglino, G. A., & Lieberoth, A. (2021). Covidistress global survey dataset on psychological and behavioural consequences of the covid-19 outbreak. *Scientific Data*, 8(1), 1–23.
- Yechiam, E., Yom-Tov, E. (2021). Unique internet search strategies of individuals with selfstated autism: Quantitative analysis of search engine users' investigative behaviors. *Journal* of Medical Internet Research, 23(7), e23829.
- Yom-Tov, E. (2019). Demographic differences in search engine use with implications for cohort selection. *Information Retrieval Journal*, 22(6), 570–580.
- Yom-Tov, E., Feraru, G., Kozdoba, M., Mannor, S., Tennenholtz, M., & Hochberg, I. (2017). Encouraging physical activity in patients with diabetes: Intervention using a reinforcement learning system. *Journal of Medical Internet Research*, 19(10), e338.
- Yom-Tov, E., Fernandez-Luque, L., Weber, I., & Crain, S. P. (2012). Pro-anorexia and pro-recovery photo sharing: A tale of two warring tribes. *Journal of Medical Internet Research*, 14(6), e151.
- Yom-Tov, E., & Lev-Ran, S. (2017). Adverse reactions associated with cannabis consumption as evident from search engine queries. JMIR Public Health and Surveillance, 3(4), e77.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

