



# Toward *Data Sense-Making* in Digital Health Communication Research: Why Theory Matters in the Age of Big Data

Edmund W. J. Lee<sup>1,2,3\*†</sup> and Andrew Z. H. Yee<sup>4†</sup>

<sup>1</sup> Department of Social and Behavioral Sciences, Harvard T. H. Chan School of Public Health, Boston, MA, United States,

<sup>2</sup> Medical Oncology, Dana–Farber Cancer Institute, Boston, MA, United States, <sup>3</sup> Wee Kim Wee School of Communication

and Information, College of Humanities, Arts, and Social Sciences, Nanyang Technological University, Singapore, Singapore,

<sup>4</sup> Humanities, Arts, and Social Sciences, Singapore University of Technology and Design, Singapore, Singapore

## OPEN ACCESS

### Edited by:

Sunny Jung Kim,  
Virginia Commonwealth University,  
United States

### Reviewed by:

Peter Johannes Schulz,  
University of Lugano, Switzerland  
Victoria Team,  
Monash University, Australia

### \*Correspondence:

Edmund W. J. Lee  
edmundlee@ntu.edu.sg

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Health Communication,  
a section of the journal  
Frontiers in Communication

**Received:** 15 October 2019

**Accepted:** 05 February 2020

**Published:** 27 February 2020

### Citation:

Lee EWJ and Yee AZH (2020) Toward  
Data Sense-Making in Digital Health  
Communication Research: Why  
Theory Matters in the Age of Big Data.  
Front. Commun. 5:11.  
doi: 10.3389/fcomm.2020.00011

The rapidly increasing volume of health data generated from digital technologies have ushered in an unprecedented opportunity for health research. Despite their promises, big data approaches in understanding human behavior often do not consider conceptual premises that provide meaning to social and behavioral data. In this paper, we update the definition of big data, and review different types and sources of health data that researchers need to grapple with. We highlight three problems in big data approaches—data deluge, data hubris, and data opacity—that are associated with the blind use of computational analysis. Finally, we lay out the importance of cultivating *health data sense-making*—the ability to integrate theory-led and data-driven approaches to process different types of health data and translating findings into tangible health outcomes—and illustrate how theorizing can matter in the age of big data.

**Keywords:** big data, artificial intelligence, machine learning, digital health, social media, wearables, patient portals, communication theory

The rapidly increasing volume of health data generated from digital health technologies such as social media, search engines, smartphones, and wearable gadgets (wearables), as well as electronic health records (EHRs) and web patient portals have ushered in an unprecedented opportunity for health communication researchers to tap on both “naturally occurring” and structured data (e.g., surveys) to improve health. To date, researchers have utilized digital data in various ways, such as predicting the onset of mental and physical illnesses (Torous et al., 2015; Merchant et al., 2019), understanding public sentiments toward health issues and patterns of information diffusion (Himmelboim and Han, 2014; Sedrak et al., 2016), as well as anticipating and managing outbreak of infectious diseases (Charles-Smith et al., 2015; Huang et al., 2016). Despite the promises and sophistication, the allure of big data approaches in understanding human behavior often do not consider conceptual premises that provide meaning to social and behavioral data (Coveney et al., 2016).

In this paper, we provide a critical review of the use of big data in health communication research<sup>1</sup>. Synthesizing concepts and research from a variety of disciplines, we contend that simply

<sup>1</sup>According to Grant and Booth (2009), a critical review is aimed at synthesizing material from a variety of sources in order to achieve conceptual innovation through the analysis of diverse material. Our goal is for this paper to serve as a starting point for discussion, and a foundation for future systematic and empirical approaches in examining health data sense-making.

using big data in research is insufficient in furthering health communication as a discipline, or to elicit insights which can improve societal health outcomes. In order to fully leverage digital health data to improve health outcomes, it is essential for health communication and public health researchers to cultivate **health data sense-making**, which is the competence to integrate theory-led and data-driven approaches to process different forms of health data. Researchers who strive for this ideal would also translate findings from big data research into tangible health outcomes.

Our objectives are 3-fold. First, we briefly define what health data sense-making is, and highlight the types of big data that health communication and public health scholars need to be acquainted with. Second, we introduce the *trinitarian* problems of health big data—(i) data deluge, (ii) data hubris, and (iii) data opacity—that are significant barriers to health researchers. Third, we explain why health data-sense making is an essential quality for health researchers who are working with—or who are intending to work with—big data.

## WHAT IS HEALTH DATA SENSE-MAKING?

Health data sense-making is an important quality that health communication researchers should possess when working with health big data, or computationally intensive projects. As described earlier, health data sense-making is the efficacy of health researchers to achieve a reasonable balance in straddling between the two worlds of *a-priori* theory building research and data-driven work. But what really constitutes engaging in health data sense-making? We postulate that health researchers must fulfill two key components. First, to be able to claim that one is engaging in health data sense-making, extending health communication theories and elucidating communication phenomenon must be at the heart of the research, and not just a peripheral afterthought. Second, there must be adoption of novel data collection or analytical methodologies in the theory-building work. For instance, while developing algorithms to effectively mine data from EHRs to detect and manage high-cost and high-risks patients is novel methodologically (e.g., Bates et al., 2014), it would not strictly constitute as health data sense-making based on our definition as health communication theory building was not the focal point of the research. Likewise, working on social media data alone would not automatically mean that the health researchers are engaged in health data sense-making. That is unless the researchers (e.g., Himelboim and Han, 2014; Kim et al., 2016) explicitly highlight how their use of computational approaches will significantly advance existing theories (e.g., social network structures, information diffusion) or our understanding of health communication behaviors.

## WHAT IS “HEALTH DATA” AND “BIG DATA” IN DIGITAL HEALTH TECHNOLOGIES?

Health data according to the General Data Protection Regulation (GDPR) in the European Union, are defined as information that are related to either the physical or mental health of a

person, or the provision of health services to an individual (GDPR Register, 2018). They could be obtained from a variety of sources such as EHRs, electronic patient portals, pharmacy records, wearables and smartphone apps, population health surveys, as well as social media. Despite the popularity of the term “big data” and the hype of what they could do to improve health, the definition is nebulous and elusive (Boyd and Crawford, 2012; Bansal et al., 2016). Yet, researchers have largely agreed that big data in health contexts possess five key characteristics—*volume*, *velocity*, *variety*, *veracity*, and *value* (Wang et al., 2018).

### Volume and Velocity

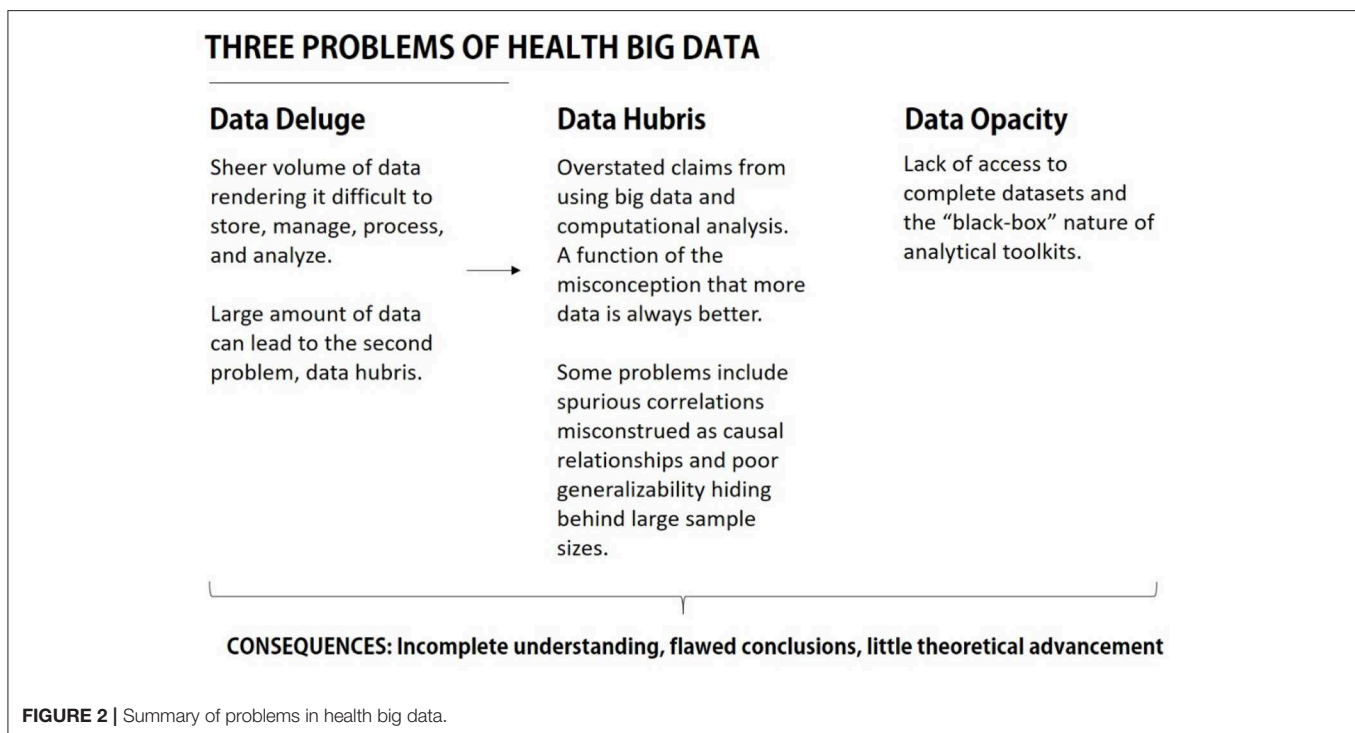
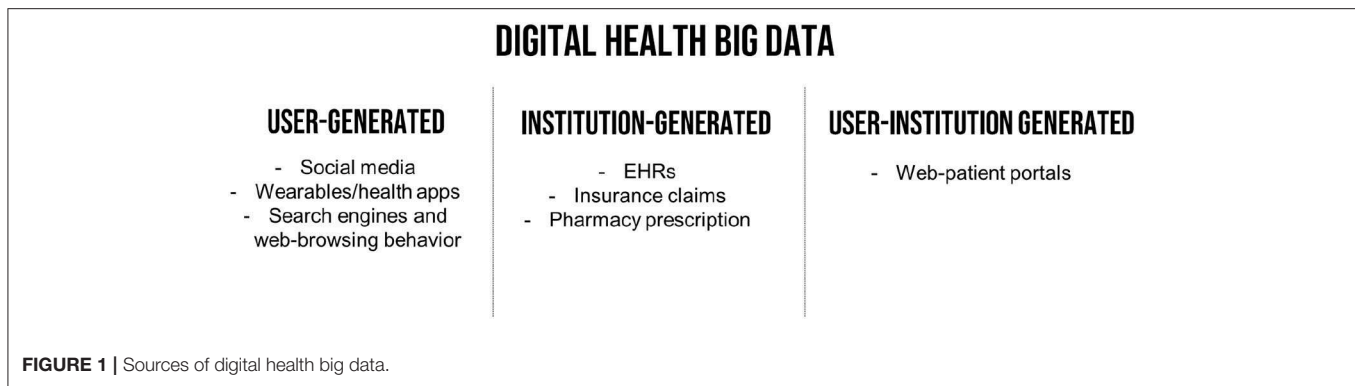
At the fundamental level, one quality of big data is the sheer volume, often generated through digital and electronic mediums such as posts on social media, location tracking on smartphones, or even collective health records of patients in hospitals. However, beyond the size of data, they have the characteristics of being generated and analyzed at an exponential rate (i.e., velocity) as compared to traditional means of data collection such as national census or health surveys. For instance, Google processes approximately 40,000 search queries every second, which equates to 3.5 billion searches daily (Internet Live Stats, 2019).

### Variety: Dimensions of Big Data

In terms of data structure, health big data consist of different *varieties* (see **Figure 1**) and could be *user-generated*, *institutional-generated*, or be a hybrid of being generated by both users and institutions (i.e., *user-institutional generated*). User-generated data are naturally occurring digital traces (Peng et al., 2019) from (a) social media (e.g., Ayers et al., 2016), (b) wearables and health apps (Casselmann et al., 2017), (c) search engines and web browsing behaviors (Mavragani et al., 2018). Institutional-generated health big data comprise of (a) EHRs which stores all the medical history of an individual, (b) claims data from insurance companies, as well as (c) pharmacy prescriptions (Wallace et al., 2014). In terms of hybrid user-institutional generated big data, an example would be web patient portals, where patients could access their medical record, interact with their healthcare providers through direct messaging, and manage their medical health such as prescription refills, schedule appointments, or accessing education content (Wells et al., 2015; Antonio et al., 2019).

### Veracity and Value

Veracity refers to the *trustworthiness* of the data. For instance, data on geolocations obtained through smartphone apps would be more accurate in giving information on people’s mobility patterns compared to self-reports. The final dimension of big data is the *value* they bring to aid decision-making, especially in the context of improving health outcomes for different population subgroups (Asokan and Asokan, 2016; Zhang et al., 2017), or enabling radiologists to detect cancer tumors more effectively (Bi et al., 2019).



## TRINITARIAN PROBLEMS OF HEALTH BIG DATA

The increasing complexity surrounding the types of big data that health communication and public health scholars work with poses significant challenges. Remaining blind, or unaware, of some of the vexing problems of health big data could lead to potential pitfalls in public health communication research design, analysis, and poor interpretation. This would dilute the promise of big data for informing health communication and public health practice. Among the many challenges, researchers should be cognizant of the trinitarian problems of health big data—(a) data deluge, (b) data hubris, and (c) data opacity (see **Figure 2**).

### The Problem of Data Deluge

Data deluge refers to the phenomenon where the sheer amount of health data being produced at such granularity

and exponentiality would be overwhelming for health communication and public health researchers to store, manage, process, and analyze. To put scale of the problem in context, it was estimated that volume of data generated by the healthcare system in the U.S. amounted to about 500 petabytes (PB)—or 1,015 bytes—in 2012, equivalent to having 10 billion file cabinets (Pramanik et al., 2017). Conservatively, this figure is expected to increase to an astounding 25,000 PB by 2020. Besides data generated via healthcare systems, the problem of data deluge is in-part driven by revolutions in different digital health technologies, computing prowess to manage data at-scale efficiently, sophisticated data storage and management systems, and the widespread adoption of data-producing technologies by the general public (Viswanath et al., 2012).

For instance, smartphone penetration in the U.S. alone rose from 35% in 2011 to 81% in 2019 (Pew Research Center, 2019a). In terms of social media use, the percentages of American adults

who use social media rose from 5% in 2005 to 72% in 2019 (Pew Research Center, 2019b). Despite having numerous scandals surrounding data privacy violation and political manipulation such as the infamous Cambridge Analytica saga, Facebook's growth remains unparalleled, with 2.4 billion monthly active users as of June 2019 (Noyes, 2019).

### Artificial Intelligence as a Solution to the Data Deluge

To cope with the huge amount of data, researchers have turned to artificial intelligence (AI) algorithms (Gupta et al., 2016; Wahl et al., 2018) to mine for insights. A key benefit of AI algorithms is their ability to computationally crunch large number of factors or variables in both interactive linear and non-linear ways to detect patterns in data (Kreatsoulas and Subramanian, 2018). While researchers have recognized that there are areas in the problem of data deluge could be effectively dealt with by AI (Maddox et al., 2019), such as identifying complex patterns in clinical settings in the area of cancer tumors detection (Bi et al., 2019), its efficacy is limited in other areas, especially in understanding motivations behind why people adhere to or reject certain health behaviors (e.g., vaccination). A blind and unassuming faith in health big data, or AI to make sense of it, would contribute to the problem of *data hubris*.

### The Problem of Data Hubris

Data hubris refers to overstated claims that arise from big data analysis. It is a consequence of an implicit adherence to radical empiricism, the belief that inductive pattern recognition from big data can substitute and perhaps overtake the traditional hypothetico-deductive model of science (Kitchin, 2014; Lazer et al., 2014). Proponents of such an approach believe in supplanting theory-driven research with data-driven insights. Anderson (2008) exemplified such a position with the claim that "we can analyze the data without hypotheses about what it might show... throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot." Specifically, data hubris is most problematic when scholars use big data incorrectly to make causal claims from an inherently inductive method of pattern recognition (Boyd and Crawford, 2012).

### Causal Inferences in Health Communication Theorizing

The issue of overstating causal relationships from big data is particularly important for health communication researchers. Although health communication is a broad and multidisciplinary field, a central goal of health communication research is to understand how the process of communication influences health outcomes (Kreps, 2001; Beck et al., 2004; Schiavo, 2013). This means that an underlying and unifying goal of health communication research is in identifying causal communication-related mechanisms that have an effect on individual's health-related outcomes and behaviors. In other words, causal inferences are fundamental in health communication research.

Undergirding various causal claims in health communication are theories that specify how and why health-related behaviors

and health outcomes are enacted (e.g., Janz and Becker, 1984; Bandura, 2004; Ryan and Deci, 2007; Goldenberg and Arndt, 2008; Rimal, 2008; Fishbein and Ajzen, 2010). Solid theoretical grounding allows for robust predictions of health behaviors, leading to more confident empirically based recommendations for policymakers and healthcare practitioners to effect real change. While big data offer an opportunity to examine health communication phenomena on an unprecedented scale, using them without theoretical considerations to guide analyses can be highly problematic. This is particularly pertinent to health communication research, as one issue we face in our field is the lack of substantial theoretical development (Beck et al., 2004; Sandberg et al., 2017). Inductive approaches to big data analysis can exacerbate this.

### Spurious Correlations

Specifically, purely data-driven big data analyses have the potential to generate large amounts of spurious findings. Spurious correlations refer to variables being associated with each other, either entirely from chance, or due to confounding factors. The enormous amount of data available in big data analyses mean that spurious correlations are more likely to be encountered (Calude and Longo, 2017). While not exactly digital traces, researchers who scanned a dead salmon in an fMRI machine while exposing it to photos of humans in a variety of social situations, found that 16 out of 8,064 brain regions in the dead salmon had a statistically significant response to the images (Bennett et al., 2009). The sheer amount of data in an fMRI scan meant that impossible findings can also be flagged out as significant. As big data datasets could be exponentially larger than an fMRI scan, spurious correlations are even more likely to be identified through various analyses. Instead of producing robust knowledge, data-driven approaches could potentially lead to a deluge of unimportant associations being flagged out. In other words, patterns derived from big data analysis are far from reliable, and causal claims arising from such findings can be nothing more than hubris. For example, the increase in Google search frequency of disease-related words during disease outbreaks is more likely due to extensive media coverage, rather than actual disease outbreak (Cervellin et al., 2017). After all, social media and search engines have the propensity to amplify risks (Strekalova, 2017; Strekalova and Krieger, 2017). Adopting the social amplification of risk framework (SARF), Strekalova and Krieger (2017) highlighted that health risks could be heightened via social media or online platforms through the mechanisms of (a) *user engagement*: where content of health risks are freely generated and diffused among social networks, (b) *media richness* (e.g., vivid descriptions or photographs depicting health risks), and (c) *signal sharing*: online communities adopting certain words, phrases, and pictures to create shared socially constructed meaning (e.g., posting selfies of people with mustache during November as part of the *Movember* campaign which raises awareness for prostate cancer) and experiences of health risks (Jacobson and Mascaro, 2016). In a study examining people's response to Centers for Disease Control and Prevention's (CDC) Facebook page, Strekalova (2017) found that even though the CDC had fewer posts of Ebola as compared to posts on health

promotion posts, the former received more attention from Facebook users.

### Misconceptions About the Generalizability of Big Data

A second problematic issue is the assumption that the enormous number of observations in big data analysis leads to generalizable findings. In contrast, big data have significant issues pertaining to sampling bias and representativeness. Specifically, existing big data studies in health research tend to utilize one platform, such as Twitter, Facebook, or EHRs (Hargittai, 2015). These different platforms have sample bias problems, with race, gender, and class being factors that determine their use (Tufekci, 2014; Hargittai, 2015). For example, only 40% of patients in the US have their information in EHRs, while gender, race, and income predict Facebook, LinkedIn, and Twitter use (Kaplan et al., 2014; Hargittai, 2015). This means that findings and conclusions drawn from single-platform big data studies are limited to the type of user that is characteristic to each platform. To add to this, supposedly random samples drawn from Twitter's Application Programming Interfaces (APIs) are sometimes skewed by the presence of sample cheaters, corporate spammers, and frequency bots (Pfeffer et al., 2018).

Relatedly, self-selection bias is particularly prevalent in social media studies that utilize hashtags. By selecting only hashtagged tweets or posts to answer certain research questions, it excludes other related content. This self-selected sample of content might lead to systematic errors which can severely impact the validity of findings (Tufekci, 2014). In sentiment analysis, hashtagged content might reflect more polarizing opinions, as in the case of the #himtoo hashtag during Brett Kavanaugh's supreme court confirmation, or the #vaccinescauseautism and #vaccineskill hashtags among the antivax population. It is possible that users who utilize hashtags have a stronger opinion on certain issues, while content produced by users who are neutral are ignored by studies that draw their data from hashtags.

### The Problem of Data Opacity

Data opacity refers to the "black-box" nature of data acquisition and analysis that health communication and public health researchers need to contend with. In terms of data acquisition and collection, public health communication researchers are often at the mercy of both large private or public corporations that have monopolistic claim on the data produced in their platforms, and work with a certain degree of ambiguity and trust with the data they are provided by these organizations, and play by their rules from the get-go. In other words, *data owners* such as social media companies, or even healthcare facilities such as hospitals, decide what data health researchers would be able to access. For instance, researchers are only able to access one percent of tweets at Twitter's API for free or at most 10% of all tweets if they have funding (Pfeffer et al., 2018). Beyond the limitations on the number of tweets, the problem of opacity is exacerbated by claims that Twitter is politically biased and censors conservatives and right-wing ideologies (Kang and Frenkel, 2018). While not conclusive, this may pose a significant problem if researchers want to investigate what people are talking

about when it comes to health issues that are closely related with political affiliations, such as abortion, gun control, tobacco use, and health care reforms.

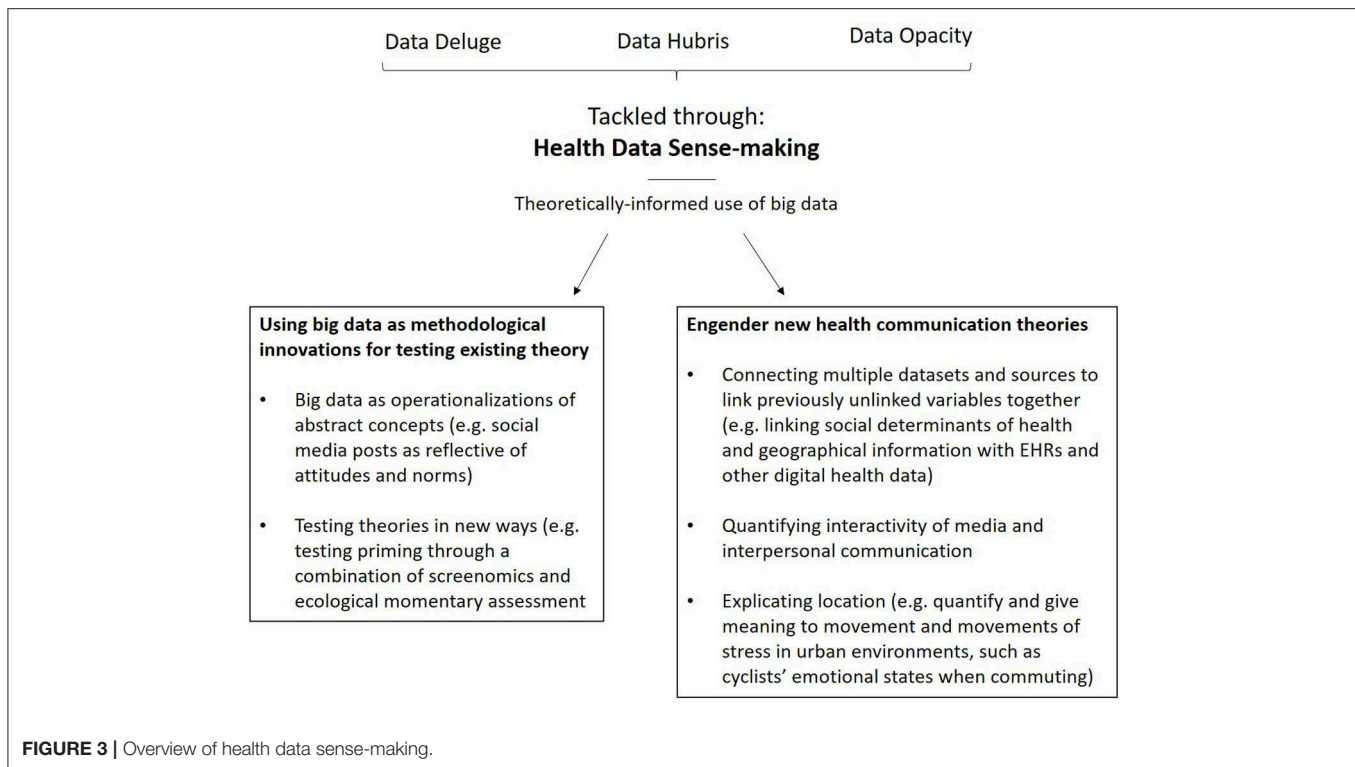
### Corporate Gatekeeping of Data

The problem of data opacity in data acquisition is not unique to social media platforms. Researchers utilizing mobile phones for health communication and public health also face the problem of data opacity. Like social media companies, researchers often do not have access to large-scale mobility data from smartphones, and would rely on data provided by telecommunication companies which often operate within strict regulations (Wesolowski et al., 2016). While researchers may be able to access large scale mobility data from telecommunication companies from open challenges such as *Data for Refugees Turkey*—an initiative by Turk Telekom (2018) to release large anonymized mobile datasets to researchers for utilizing big data solutions to improve living conditions for Syrian refugees—researchers ultimately would not have full details as to how the data was prepared, and if certain details were left out.

Data opacity at the data collection level poses practical challenges for health researchers who rely on them for their research. Recently, Facebook partnered with Harvard University and the Social Science Research Council to forge an industry-academic partnership, and pledged to support academic scholars to use a subset of Facebook data to study the impact of social media on democracy (Social Science One, 2018). The collaboration would involve Facebook providing URLs to researchers who have been granted ethical approval from their respective Institutional Review Boards (IRB) for their research. Yet, due to Facebook being unable to provide what was promised in the original request for proposals in 2018 (King and Persily, 2019), researchers would need to work with reduced versions of the data until the full release. While this is not in the health context, it shows the extent of reliance that health researchers have on companies who play a gate-keeping role in health big data.

### Analytical Tools as Black Boxes

In addition to data acquisition, data opacity occurs at the level of analysis as well. For instance, for health researchers analyzing social media data, one very common algorithm to deploy is topic modeling. Topic modeling is a text mining approach for identifying "themes" computationally in a large corpus of unstructured texts (Richardson et al., 2014). One of the most widely used algorithms for topic modeling is Latent Dirichlet Allocation (LDA), which is a popular technique used by researchers to examine what people are talking about online pertaining to health (Bian et al., 2017). In using LDA, researchers need to specify *a priori* how many latent topics there are. For deciding on the final number of topics there are in a given text, researchers typically turn to computational measures such as *perplexity scores* or the likes—a measure of goodness of fit, where lower scores meant that the number of topics fit the LDA model (Blei et al., 2003; Fung et al., 2017). The strict reliance on computational power ignores nuanced meaning in languages that



may skew the results, or how certain geo-political environmental contexts may augment the nature of the topics.

A second example of data opacity in the context of data analysis is in the use of geospatial analytics on location or physiological data obtained from smartphones or wearables. While some would argue that mobile sensing in public health has a higher degree of objectivity in terms of the data it could capture (Chaix, 2018)—GPS location would inform researchers where individuals have been—yet, there are multiple factors at the analysis stage that dilute this objectivity. For instance, in analyzing proximity to key features (e.g., hospitals, services), a common technique is *buffer analysis*, which allows researchers to create artificial boundaries around the features of interests and count how many data points (e.g., mobility patterns) are within the artificial boundaries (ArcMap, 2019). For instance, research in understanding opioid overdose has used buffer analysis to identify the total number of discarded needles in a given area (Bearnot et al., 2018). The selection of such boundaries for buffer analysis is inherently subjective, and the assumptions of such selection would need to be supported theoretically.

## TOWARD HEALTH DATA SENSE-MAKING

Our proposed solution to the trifecta of problems—data deluge, hubris, and opacity—is for researchers to cultivate health data sense-making. To do that, we need to adopt approaches to big data research which prioritize the use of theory in providing meaning to data. This is not a new idea, and is fundamental to the hypothetico-deductive model of science (Godfrey-Smith, 2003).

There are two ways in which health researchers can effectively practice health data sense-making. First, health researchers should view big data primarily as methodological innovations which can lend itself to testing existing health communication theory in novel ways; and second, health researchers should utilize these methodological innovations to engender new health communication theory that is made possible by the nature of big data (see Figure 3).

### Viewing Big Data as Methodological Innovations for the Testing of Existing Theories

The most immediate way to cultivate health data sense-making is for researchers to utilize the possibilities afforded by big data to develop innovative methods for the testing of existing theory. To do this, health researchers must approach research projects involving big data from a theory-driven, rather than data-driven, perspective. As described by Chaffee (2009), theory development should come from careful explication of abstract concepts. This involves both logically explaining the theoretical underpinnings behind hypothesized effects between abstract concepts and conducting careful explication and operationalization of those abstract concepts for formal testing.

### Big Data as Operationalizations of Abstract Theoretical Concepts

While existing research in health communication research utilize surveys and experiments to measure abstract concepts such as attitudes and beliefs, it might be possible to view social

media content as operationalizations of abstract concepts as well. Specifically, social media posts can be operationalizations of sentiment, or attitude, toward certain health behaviors. For example, in the context of alcohol consumption, production of social media content—tweets or posts—can be conceptualized as an underlying positive or negative attitude toward drinking, while the consumption of social media content (what people are exposed to) regarding drinking can be conceptualized as social norms (Cavazos-Rehg et al., 2015). In combination with other measures, these can be used as measures within existing theoretical frameworks such as the theory of normative social behavior or the theory of planned behavior (Ajzen, 1991; Rimal and Lapinski, 2015).

Despite this, formal explication of how existing theoretical constructs such as attitude and beliefs can be measured through big data is rare. In one study, researchers looked at how both visual and textual posts about alcohol on Facebook predicts binge drinking, in addition to attitude and injunctive norms (D'Angelo et al., 2014). Instead of viewing the social media posts as a distinct and separate construct to attitude in a predictive model, we propose that health communication researchers should aim to explicate what these specific data are reflecting. Could they be reflective of an individuals' attitude toward drinking, or something else? Health communication researchers must begin clarifying what specific pieces of big data mean.

### Testing Theories in New Ways

Beyond the process of concept explication, big data also offer opportunities for the testing of existing theories that was previously impossible. For example, traditional ways of testing priming effects utilize laboratory experiments where researchers manipulate stimuli of interest (e.g., Alhabash et al., 2016). Instead of testing priming effects in a lab, big data tools allow us to examine these effects in a naturalistic setting.

Take for example *screenomics*, a novel approach in measuring digital media use through the capturing of screenshots periodically (e.g., screenshot every 5 s), before being encrypted and sent to university servers for processing, in which textual and image data are extracted using computer vision technology and natural language processing (Reeves et al., 2019). Instead of manipulating Facebook alcohol ads to investigate priming effects intention to consume alcohol, as Alhabash et al. (2016) had done, *screenomics* paired with ecological momentary assessment (EMA) measuring alcohol consumption intentions can provide us with a unique way of investigating priming effects naturalistically. Furthermore, with the sheer amount of data generated, in addition to examining priming effects across participants, it is possible to examine priming effects within an individual (e.g., whether exposure to alcohol images and posts primes a *specific* individual's intention to consume alcohol) by using a person's daily screenshot data and EMA responses as a single data point. This leads to the potential testing of health communication theories through *both* an idiographic and nomothetic approach. Such naturalistic ways of collecting media and behavioral data, addresses one of the biggest criticisms of traditional experimental media effects research—the artificiality of the environment in which stimuli are manipulated, and the artificiality of the stimuli themselves (Livingstone, 1996).

## Engender New Health Communication Theories Through Health Data Sense-Making

In addition to enhancing existing health communication theories, health data sense-making could potentially engender new health communication theories and concepts through the synergistic and iterative combination of traditional *a-priori* social scientific communication theorizing and data-driven approaches. When health researchers pair health communication theorizing with computational prowess, it avoids the reckless use of algorithms that might lead to spurious correlations, or engagement in *p-hacking*, where researchers selectively report positive results by running computational models many times to get statistical significance (Head et al., 2015). After all, the key word in data science is the practice of *science*, and big data and AI techniques are the computational enablers for large-scale and in-depth theory testing through an iterative process. While the potential for engendering new health communication theories and models is limitless, we list three areas how health data sense-making could be useful for developing new health communication theories and conceptual explication.

### Connecting the Dots for Health Disparities

First, as health communication scholars increasingly need to work with big data from diverse sources, there is a potential to connect the dots by drawing upon diverse datasets on how social determinants exacerbate health disparities (Lee and Viswanath, 2020). There are two key developments that would enable health data sense-making for the context of health disparities. First, the incorporation of factors such as social determinants and geographical variables into existing EHRs (Zhang et al., 2017). The second is the move toward utilizing digital and traditional media and social norms in nudging individuals to take responsibilities of their own health through active health monitoring (Patel et al., 2015; Hentschel et al., 2016; Ho et al., 2016).

The potential to link social determinants, demographic, and geographical variables with how people are utilizing digital devices for health and subsequent health outcomes would shed light on how different forms of *communication inequalities*—differences among social groups in generation and use of information that contribute to differential health outcomes—serve as intermediary mechanisms that moderate the effects of social determinants on health disparities (Viswanath et al., 2012; Lee and Viswanath, 2020). The ability to synthesize insights from datasets would contribute to macro health communication theorizing (e.g., ecological health model, social cognitive model), and enable scholars to specifically tease out how external and individual determinants factors collectively influence health outcomes and behaviors (Bandura, 2001; Richard et al., 2011; Lee et al., 2016; Lee and Viswanath, 2020).

### Quantifying Interactivity of Media and Interpersonal Influence

The second way new health communication theories could emerge is in the area of quantifying the interactivity of media consumption and interpersonal networks on health communication behaviors. *Screenomics*, for example, has the

potential to allow health communication researchers to explicate *health information seeking* beyond traditional survey measures, and specifically examine the process of health information seeking, and how different groups seek health information (Reeves et al., 2019). In terms of quantifying the interactivity of media and network effects, researchers could test their *a-priori* hypotheses on how they expect information flow and interpersonal networks to influence health outcomes through the use of dynamic or agent-based modeling (Zhang et al., 2015; Song and Boomgaarden, 2017), and compare how *reinforcing spirals* (Slater, 2007) of media effects and interpersonal influence on health behaviors differ in a simulated environment and in real environment.

### Explicating Location—Modeling the Ebb and Flow of *in situ* Health Behaviors

The third area which there is a potential for emergence of new health communication theories is in the area of explicating different dimensions of locations within health communication theories and examine the variability in health outcomes at a granular level. Traditionally, health communication theories such as the *structural influence model of communication* postulates that social determinants such as geographical locations in terms of neighborhood and urbanicity (Kontos and Viswanath, 2011) are underlying macro-level factors influencing health outcomes. Yet, how scholars conceptualize locations are often based on artificial administrative boundaries (e.g., census tracts or zip codes), which would be problematic as people are highly mobile and not constrained to a single location.

The advancement in collection of location data, together with physiological and behavioral data through mobile health apps and wearables, would enable researchers to do a deep dive into how specific locations matter in health outcomes. For instance, by drawing upon geolocations, galvanic skin response, skin temperature, and heart rate variability, researchers could identify areas quantify *moments* and *movements of stress* in an urban environment, such as cyclists' emotional states when traveling in certain paths (Kyriakou et al., 2019). In other words, coupled with geospatial analytical tools such as emerging hotspot

analysis, researchers could accurately identify potential hotspots where traffic accidents are likely to happen because of the stress-clustering. In addition, by utilizing geolocation tracking from smartphones, researchers would be conceptualize and concretize *mobility signatures*, and connect how real-time exposure to point-of-sale tobacco marketing, differential tobacco product pricing, location of tobacco retail outlets, influence quit attempts (Kirchner et al., 2012).

## CONCLUSION

The advent of health big data is both a boon and bane to researchers. The ability to leverage on big data for health communication depends largely on researchers' ability to see beyond the hype, separate wheat from the chaff, and put the horse before the cart. Otherwise, the blind pursuit and application of AI and big data would simply lead to artificial inflation of findings. Health data-sense making—the science and art of prioritizing theories before computation—would place health researchers in good stead to avoid the pitfalls of health big data, and move the field of health communication forward with new perspectives on improving health behaviors by synthesizing health data from individuals, their interaction with environmental factors, significant others, and the media. After all, health data sense-making is a critical response to the call to *humanize* data instead of *data-fying* humans (Israni and Verghese, 2018), which would be absolutely critical for moving the field of health communication forward in the digital age.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

The article processing fee was funded by Singapore University of Technology and Design's Faculty Early Career Award (RGFECA19004).

## REFERENCES

- Ajzen, I. (1991). The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* 50, 179–211. doi: 10.1016/0749-5978(91)90020-T
- Alhabash, S., McAlister, A. R., Kim, W., Lou, C., Cunningham, C., Quilliam, E. T., et al. (2016). Saw it on Facebook, drank it at the bar! Effects of exposure to Facebook alcohol ads on alcohol-related behaviors. *J. Interact. Advertis.* 16, 44–58. doi: 10.1080/15252019.2016.1160330
- Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *WIRED*. Retrieved from: <https://www.wired.com/2008/06/pb-theory/>
- Antonio, M. G., Petrovskaya, O., and Lau, F. (2019). Is research on patient portals attuned to health equity? A scoping review. *J. Am. Med. Inform. Assoc.* 26, 871–883. doi: 10.1093/jamia/ocz054
- ArcMap (2019). *How Buffer (Analysis) Works*. Available online at: <http://desktop.arcgis.com/en/arcmap/10.3/tools/analysis-toolbox/how-buffer-analysis-works.htm> (accessed January 3, 2020).
- Asokan, G. V., and Asokan, V. (2016). Leveraging “big data” to enhance the effectiveness of “one health” in an era of health informatics. *J. Epidemiol. Glob. Health* 5, 311–314. doi: 10.1016/j.jegh.2015.02.001
- Ayers, J. W., Westmaas, J. L., Leas, E. C., Benton, A., Chen, Y., Dredze, M., et al. (2016). Leveraging big data to improve health awareness campaigns: a novel evaluation of the great American smokeout. *JMIR Public Health Surveil.* 2:e16. doi: 10.2196/publichealth.5304
- Bandura, A. (2001). Social cognitive theory of mass communication. *Media Psychol.* 3, 265–299. doi: 10.1207/S1532785XMEP0303\_03
- Bandura, A. (2004). Health promotion by social cognitive means. *Health Educ. Behav.* 31, 143–164. doi: 10.1177/1090198104263660
- Bansal, S., Chowell, G., Simonsen, L., Vespignani, A., and Viboud, C. (2016). Big data for infectious disease surveillance and modeling. *J. Infect. Dis.* 214, S375–S379. doi: 10.1093/infdis/jiw400
- Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., and Escobar, G. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff.* 33, 1123–1131. doi: 10.1377/hlthaff.2014.0041



- Bearnot, B., Pearson, J. F., and Rodriguez, J. A. (2018). Using publicly available data to understand the opioid overdose epidemic: geospatial distribution of discarded needles in Boston, Massachusetts. *Am. J. Public Health* 108, 1355–1357. doi: 10.2105/AJPH.2018.304583
- Beck, C. S., Benitez, J. L., Edwards, A., Olson, A., Pai, A., and Torres, M. B. (2004). Enacting “health communication”: the field of health communication as constructed through publication in scholarly journals. *Health Commun.* 16, 475–492. doi: 10.1207/s15327027hc1604\_5
- Bennett, C. M., Baird, A. A., Miller, M. B., and Wolford, G. L. (2009). “Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: an argument for proper multiple comparisons correction,” in *15th Annual Meeting of the Organization for Human Brain Mapping* (San Francisco, CA).
- Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., et al. (2019). Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J. Clin.* 69, 127–157. doi: 10.3322/caac.21552
- Bian, J., Zhao, Y., Salloum, R. G., Guo, Y., and Wang, M. (2017). Using social media data to understand the impact of promotional information on laypeople’s discussions: a case study of lynch syndrome. *J. Med. Internet Res.* 19, 1–16. doi: 10.2196/jmir.9266
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022.
- Boyd, D., and Crawford, K. (2012). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Inform. Commun. Soc.* 15, 662–679. doi: 10.1080/1369118X.2012.678878
- Calude, C. S., and Longo, G. (2017). The deluge of spurious correlations in big data. *Found. Sci.* 22, 595–612. doi: 10.1007/s10699-016-9489-4
- Casselmann, J., Onopa, N., and Khansa, L. (2017). Wearable healthcare: lessons from the past and a peek into the future. *Telemat. Informat.* 34, 1011–1023. doi: 10.1016/j.tele.2017.04.011
- Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S. J., and Bierut, L. J. (2015). “Hey everyone, I’m drunk.” An evaluation of drinking-related Twitter chatter. *J. Stud. Alcohol Drugs* 76, 635–639. doi: 10.15288/jsad.2015.76.635
- Cervellini, G., Comelli, I., and Lippi, G. (2017). Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J. Epidemiol. Glob. Health* 7, 185–189. doi: 10.1016/j.jegh.2017.06.001
- Chaffee, S. H. (2009). “Thinking about theory,” in *An Integrated Approach to Communication Theory and Research*, 2nd Edn, eds M. B. Salwen and D. W. Stacks (Mahwah, NJ: Lawrence Erlbaum Associates), 12–29.
- Chaix, B. (2018). Mobile sensing in environmental health and neighborhood research. *Annu. Rev. Public Health* 39, 367–384. doi: 10.1146/annurev-publhealth-040617-013731
- Charles-Smith, L. E., Reynolds, T. L., Cameron, M. A., Conway, M., Eric, H., Lau, Y., et al. (2015). Using social media for actionable disease surveillance and outbreak Management: a systematic literature review. *PLoS ONE* 10:e0139701. doi: 10.1371/journal.pone.0139701
- Coveney, P. V., Dougherty, E. R., and Highfield, R. R. (2016). Big data need big theory too. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 374:20160153. doi: 10.1098/rsta.2016.0153
- D’Angelo, J., Kerr, B., and Moreno, M. A. (2014). Facebook displays as predictors of binge drinking. *Bull. Sci. Technol. Soc.* 34, 159–169. doi: 10.1177/0270467615584044
- Fishbein, M., and Ajzen, I. (2010). *Predicting and Changing Behavior: The Reasoned Action Approach*. New York, NY: Psychology Press.
- Fung, I. C. H., Jackson, A. M., Ahweyevu, J. O., Grizzle, J. H., Yin, J., Tsz, Z. H. Z., et al. (2017). #GlobalHealth Twitter conversations on #Malaria, #HIV, #TB, #NCDs, and #NTDS: A cross-sectional analysis. *Ann. Glob. Health* 83, 682–690. doi: 10.1016/j.aogh.2017.09.006
- GDPR Register (2018). *Healthcare Sector: How to Comply With GDPR?* Available online at: <https://www.gdprregister.eu/gdpr/healthcare-sector-gdpr/>
- Godfrey-Smith, P. (2003). *Theory and Reality: An Introduction to the Philosophy of Science*, 1st Edn. Chicago, IL: The University of Chicago Press.
- Goldenberg, J. L., and Arndt, J. (2008). The implications of death for health: a terror management health model for behavioral health promotion. *Psychol. Rev.* 115, 1032–1053. doi: 10.1037/a0013326
- Grant, M. J., and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info. Libr. J.* 26, 91–108. doi: 10.1111/j.1471-1842.2009.00848.x
- Gupta, P., Sharma, A., and Jindal, R. (2016). Scalable machine-learning algorithms for big data analytics: a comprehensive review. *WIREs Data Mining Knowl. Discov.* 6, 194–214. doi: 10.1002/widm.1194
- Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *Ann. Am. Acad. Political Soc. Sci.* 659, 63–76. doi: 10.1177/0002716215570866
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., and Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biol.* 13:e1002106. doi: 10.1371/journal.pbio.1002106
- Hentschel, M. A., Haaksma, M. L., and van de Belt, T. H. (2016). Wearable technology for the elderly: underutilized solutions. *Eur. Geriatr. Med.* 7, 399–401. doi: 10.1016/j.eurger.2016.07.008
- Himmelboim, I., and Han, J. Y. (2014). Cancer talk on Twitter: community structure and information sources in breast and prostate cancer social networks. *J. Health Commun.* 19, 210–225. doi: 10.1080/10810730.2013.811321
- Ho, S. S., Lee, E. W. J., Ng, K., Leong, G. S. H., and Tham, T. H. M. (2016). For fit’s sake: a norms-based approach to healthy behaviors through influence of presumed media influence. *Health Commun.* 31, 1072–1080. doi: 10.1080/10410236.2015.1038772
- Huang, D., Wang, J., Huang, J., Sui, D. Z., and Zhang, H. (2016). Towards identifying and reducing the bias of disease information extracted from search engine data. *PLoS Comput. Biol.* 12:e1004876. doi: 10.1371/journal.pcbi.1004876
- Internet Live Stats (2019). *Google Search Statistics*. Available online at: <https://www.internetlivestats.com/google-search-statistics/>
- Israni, S. T., and Verghese, A. (2018). Humanizing artificial intelligence. *JAMA* 169, 20–29. doi: 10.1001/jama.2018.19398
- Jacobson, J., and Mascaro, C. (2016). Movember: Twitter conversations of a hairy social movement. *Soc. Media Soc.* 2, 1–12. doi: 10.1177/2056305116637103
- Janz, N. K., and Becker, M. H. (1984). The health belief model: a decade later. *Health Educ. Q.* 11, 1–47. doi: 10.1177/109019818401100101
- Kang, C., and Frenkel, S. (2018). Republicans accuse Twitter of bias against conservatives. *The New York Times*. Available online at: <https://www.nytimes.com/2018/09/05/technology/lawmakers-facebook-twitter-foreign-influence-hearing.html>
- Kaplan, R. M., Chambers, D. A., and Glasgow, R. E. (2014). Big data and large sample size: a cautionary note on the potential for bias. *Clin. Transl. Sci.* 7, 342–346. doi: 10.1111/cts.12178
- Kim, E., Hou, J., Han, J. Y., and Himmelboim, I. (2016). Predicting retweeting behavior on breast cancer social networks: network and content characteristics. *J. Health Commun.* 21, 479–486. doi: 10.1080/10810730.2015.1103326
- King, G., and Persily, N. (2019). *Building Infrastructure for Studying Social Media’s Role in Elections and Democracy*. Available online at: <https://socialscience.one/blog/building-infrastructure-studying-social-media’s-role-elections-and-democracy> (accessed January 3, 2020).
- Kirchner, T. R., Vallone, D., Cantrell, J., Anesetti-Rothermel, A., Pearson, J., Cha, S., et al. (2012). “Individual mobility patterns and real-time geo-spatial exposure to point-of-sale tobacco marketing,” in *WH’12 Proceedings of the Conference on Wireless Health* (San Diego, CA), 1–8.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data Soc.* 1, 1–12. doi: 10.1177/2053951714528481
- Kontos, E. Z., and Viswanath, K. (2011). Cancer-related direct-to-consumer advertising: a critical review. *Nat. Rev. Cancer* 11, 142–150. doi: 10.1038/nrc2999
- Kreatsoulas, C., and Subramanian, S. V. (2018). Machine learning in social epidemiology: learning from experience. *SSM Popul. Health* 4, 347–349. doi: 10.1016/j.ssmph.2018.03.007
- Kreps, G. L. (2001). The evolution and advancement of health communication inquiry. *Ann. Int. Commun. Assoc.* 24, 231–253. doi: 10.1080/23808985.2001.11678988
- Kyriakou, K., Resch, B., Sagl, G., Petutschnig, A., Werner, C., Niederseer, D., et al. (2019). Detecting moments of stress from measurements of wearable physiological sensors. *Sensors* 19, 1–26. doi: 10.3390/s19173805
- Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of google flu: traps in big data analysis. *Science* 343, 1203–1205. doi: 10.1126/science.1248506
- Lee, E. W. J., Shin, M., Kawaja, A., and Ho, S. S. (2016). The augmented cognitive mediation model: examining antecedents of factual and structural breast cancer

- knowledge among Singaporean women. *J. Health Commun.* 21, 583–592. doi: 10.1080/10810730.2015.1114053
- Lee, E. W. J., and Viswanath, K. (2020). Big data in context: addressing the twin perils of data absenteeism and chauvinism in the context of health disparities. *J. Med. Internet Res.* 22, 1–7. doi: 10.2196/16377
- Livingstone, S. (1996). “On the continuing problems of media effects research,” in *Mass Media and Society, 2nd Edn*, eds J. Curran and M. Gurevitch (London), 305–324. Retrieved from: [http://eprints.lse.ac.uk/21503/1/On\\_the\\_continuing\\_problems\\_of\\_media\\_effects\\_research%28LSERO%29.pdf](http://eprints.lse.ac.uk/21503/1/On_the_continuing_problems_of_media_effects_research%28LSERO%29.pdf)
- Maddox, T. M., Rumsfeld, J. S., and Payne, P. R. O. (2019). Questions for artificial intelligence in health care. *JAMA* 321, 31–32. doi: 10.1001/jama.2018.18932
- Mavragani, A., Ochoa, G., and Tsagarakis, K. P. (2018). Assessing the methods, tools, and statistical approaches in Google trends research: systematic review. *J. Med. Internet Res.* 20, 1–20. doi: 10.2196/jmir.9366
- Merchant, R. M., Asch, D. A., Crutchley, P., Ungar, L. H., Guntuku, S. C., Eichstaedt, J. C., et al. (2019). Evaluating the predictability of medical conditions from social media posts. *PLoS ONE* 14:e0215476. doi: 10.1371/journal.pone.0215476
- Noyes, D. (2019). *The Top 20 Valuable Facebook Statistics*. Available online at: <https://zephoria.com/top-15-valuable-facebook-statistics/> (accessed January 3, 2020).
- Patel, M., Asch, D., and Volpp, K. (2015). Wearable devices as facilitators, not drivers, of health behavior change. *JAMA* 313, 459–460. doi: 10.1001/jama.2014.14781
- Peng, T. Q., Liang, H., and Zhu, J. J. H. (2019). Introducing computational social science for Asia-Pacific communication research. *Asian J. Commun.* 29, 205–216. doi: 10.1080/01292986.2019.1602911
- Pew Research Center (2019a). *Mobile Fact Sheet*. Available online at: <https://www.pewinternet.org/fact-sheet/mobile/>
- Pew Research Center (2019b). *Social Media Fact Sheet*. Available online at: <http://www.pewinternet.org/fact-sheet/social-media/>
- Pfeffer, J., Mayer, K., and Morstatter, F. (2018). Tampering with Twitter’s sample API. *EPJ Data Sci.* 7:50. doi: 10.1140/epjds/s13688-018-0178-0
- Pramanik, M. I., Lau, R. Y. K., Demirkan, H., and Azad, M. A. K. (2017). Smart health: big data enabled health paradigm within smart cities. *Expert Syst. Appl.* 87, 370–383. doi: 10.1016/j.eswa.2017.06.027
- Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., et al. (2019). Screenomics: a framework to capture and analyze personal life experiences and the ways that technology shapes them. *Hum. Comp. Interact.* 1–52. doi: 10.1080/07370024.2019.1578652
- Richard, L., Gauvin, L., and Raine, K. (2011). Ecological models revisited: their uses and evolution in health promotion over two decades. *Annu. Rev. Public Health* 32, 307–326. doi: 10.1146/annurev-publhealth-031210-101141
- Richardson, G. M., Bowers, J., Woodwill, A. J., Barr, J. R., Gawron, J. M., and Levine, R. A. (2014). Topic models: a tutorial with R. *Int. J. Semant. Comput.* 8, 85–98. doi: 10.1142/S1793351X14500044
- Rimal, R. N. (2008). Modeling the relationship between descriptive norms and behaviors: a test and extension of the theory of normative social behavior (TNSB). *Health Commun.* 23, 103–116. doi: 10.1080/10410230801967791
- Rimal, R. N., and Lapinski, M. K. (2015). A re-explication of social norms, ten years later. *Commun. Theory* 25, 393–409. doi: 10.1111/comt.12080
- Ryan, R. M., and Deci, E. L. (2007). “Active human nature: self-determination theory and the promotion and maintenance of sport, exercise, and health,” in *Intrinsic Motivation and Self-determination in Exercise and Sport*, eds M. S. Hagger and N. L. D. Chatzisarantis (Champaign, IL: Human Kinetics), 1–20.
- Sandberg, H., Fristedt, R. A., Johansson, A., and Karregard, S. (2017). Health communication an in-depth analysis of the area of expertise and research literature 2010–2016. *Eur. J. Public Health* 27(Suppl. 3):ckx186.137. doi: 10.1093/eurpub/ckx186.137
- Schiavo, R. (2013). *Health Communication: From Theory to Practice*, 2nd Edn. San Francisco, CA: Jossey-Bass.
- Sedrak, M. S., Cohen, R. B., Merchant, R. M., and Schapira, M. M. (2016). Cancer communication in the social media age. *JAMA Oncol.* 2, 822–823. doi: 10.1001/jamaoncol.2015.5475
- Slater, M. D. (2007). Reinforcing spirals: the mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Commun. Theory* 17, 281–303. doi: 10.1111/j.1468-2885.2007.00296.x
- Social Science One (2018). *Our Facebook Partnership*. Available online at: <https://socialscience.one/our-facebook-partnership>
- Song, H., and Boomgaarden, H. G. (2017). Dynamic spirals put to test: an agent-based model of reinforcing spirals between selective exposure, interpersonal networks, and attitude polarization. *J. Commun.* 67, 256–281. doi: 10.1111/jcom.12288
- Strekalova, Y. A. (2017). Health risk information engagement and amplification on social media: news about an emerging pandemic on Facebook. *Health Educ. Behav.* 44, 332–339. doi: 10.1177/1090198116660310
- Strekalova, Y. A., and Krieger, J. L. (2017). Beyond words: amplification of cancer risk communication on social media. *J. Health Commun.* 22, 849–857. doi: 10.1080/10810730.2017.1367336
- Torous, J., Staples, P., and Onnela, J. P. (2015). Realizing the potential of mobile mental health: new methods for new data in psychiatry. *Curr. Psychiatry Rep.* 17, 1–7. doi: 10.1007/s11920-015-0602-0
- Tufekci, Z. (2014). “Big questions for social media big data: Representativeness, validity and other methodological pitfalls,” in *ICWSM’14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*. Ann Arbor, MI.
- Turk Telekom (2018). *D4R*. Available online at: <https://d4r.turktelekom.com.tr>
- Viswanath, K., Nagler, R. H., Bigman-Galimore, C. A., McCauley, M. P., Jung, M., and Ramanadhan, S. (2012). The communications revolution and health inequalities in the 21st century: implications for cancer control. *Cancer Epidemiol. Biomark. Prev.* 21, 1701–1708. doi: 10.1158/1055-9965.EPI-12-0852
- Wahl, B., Cossy-gantner, A., Germann, S., and Schwalbe, N. R. (2018). Artificial intelligence (AI) and global health : how can AI contribute to health in resource-poor settings? *BMJ Glob. Health* 3, 1–7. doi: 10.1136/bmjgh-2018-000798
- Wallace, P. J., Shah, N. D., Dennen, T., Bleicher, P. A., and Crown, W. H. (2014). Optum labs: building a novel node in the learning health care system. *Health Aff.* 33, 1187–1194. doi: 10.1377/hlthaff.2014.0038
- Wang, Y., Kung, L. A., and Byrd, T. A. (2018). Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Change* 126, 3–13. doi: 10.1016/j.techfore.2015.12.019
- Wells, S., Rozenblum, R., Park, A., Dunn, M., and Bates, D. W. (2015). Organizational strategies for promoting patient and provider uptake of personal health records. *J. Am. Med. Inform. Assoc.* 22, 213–222. doi: 10.1136/amiajnl-2014-003055
- Wesolowski, A., Buckee, C. O., Engo-Monsen, K., and Metcalf, C. J. E. (2016). Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *J. Infect. Dis.* 214, S414–S420. doi: 10.1093/infdis/jiw273
- Zhang, J., Tong, L., Lamberson, P. J., Durazo-Arvizu, R. A., Luke, A., and Shoham, D. A. (2015). Leveraging social influence to address overweight and obesity using agent-based models: the role of adolescent social networks. *Soc. Sci. Med.* 125, 203–213. doi: 10.1016/j.socscimed.2014.05.049
- Zhang, X., Perez-Stable, E., Bourne, P., Pehrah, E., Duru, K., Breen, N., et al. (2017). Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. *Ethnic. Dis.* 27, 95–106. doi: 10.18865/ed.27.2.95

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Lee and Yee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.