Article

BIG DATA & SOCIETY

Big Data & Society
July–December 2018: 1–21
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2053951718816727
journals.sagepub.com/home/bds
SAGE

Conceptualizations of Big Data and their epistemological claims in healthcare: A discourse analysis

Marthe Stevens, Rik Wehrens and Antoinette de Bont

Abstract

In recent years, the healthcare field welcomed an emerging field of practices captured under the umbrella term 'Big Data'. This term is surrounded with positive rhetoric and promises about the ability to analyse real-world data quickly and comprehensively. Such rhetoric is highly consequential in shaping debates on Big Data. While the fields of Science and Technology Studies and Critical Data Studies have been instrumental in elaborating the neglected and problematic dimensions of Big Data, it remains an open question how and to what extent such insights become embedded in other fields. In this paper, we analyse the epistemological claims that accompany Big Data in the healthcare domain. We systematically searched scientific literature and selected 206 editorials as these reflect on developments in the domain. Through an interpretive analysis, we construct five ideal-typical discourses that all frame Big Data in specific ways. Three of the discourses (the modernist, instrumentalist and pragmatist) frame Big Data in positive terms and disseminate a compelling rhetoric. Metaphors of 'capturing', 'illuminating' and 'harnessing' data presume that Big Data are benign and leading to valid knowledge. The scientist and critical-interpretive discourses question the objectivity and effectivity claims of Big Data. Metaphors of 'selecting' and 'constructing' data illustrate another political message, framing Big Data as limited. We conclude that work in the critical-interpretive discourse has not broadly infiltrated the medical domain. Ways to better integrate aspects of the discourse in the healthcare domain are urgently needed.

Keywords

Big Data, evidence, healthcare, discourse analysis, systematic review, editorials

Introduction

In recent years, the healthcare field has welcomed an emerging field of practices captured under the umbrella term of 'Big Data'.[1] Big Data initiatives are welcomed because of their envisioned benefits for faster and more representative knowledge[2] that is presumed to improve the process, management and predictability of care (Murdoch and Detsky, 2013). The healthcare field traditionally favours high-quality evidence from randomized controlled trials (RCTs) and observational studies to guide treatment decisions and to organize the field (Timmermans and Berg, 2003). However, as the persistent discussions about evidence-based medicine show, the field has been struggling with the reductionist and generalized character of this evidence (Berwick, 2016; Greenhalgh et al., 2014). Patient guidelines are, for example, often based on time-consuming RCTs and done on selective populations, which makes it hard to extrapolate results to individual patients (Felder and Meerding, 2017). Big Data seem to offer an attractive alternative and are surrounded by claims of quick and comprehensive analysis of data and 'with the aura of truth, objectivity and accuracy' (Boyd and Crawford, 2012: 663). These grand promises lead to a positive rhetoric that surrounds the term and that drives implementation of Big Data in healthcare.

Department of Health Care Governance, Erasmus School of Health Policy & Management, the Netherlands

Corresponding author:
Marthe Stevens, Department of Health Care Governance, Erasmus School of Health Policy & Management, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands.
Email: stevens@eshpm.eur.nl

Publications about Big Data frequently discuss topics related to knowledge generation, evidence and causation (e.g. Anderson, 2008; Mayer-Schönberger and Cukier, 2014). Provocatively, these publications celebrate the inevitable decline of traditional research as Big Data are supposed to handle large volumes of messy real-world data more efficiently and can uncover hidden correlations. In response to these claims, there has been a recurrent call for more studies into the epistemological implications of Big Data (Boyd and Crawford, 2012; Crawford et al., 2014; Mittelstadt and Floridi, 2016), which scholars have started to address. As a result, a critical scholarly discourse that reflects on how Big Data shape our knowledge and understanding is forming in, primarily, the fields of Science and Technology Studies (STS) and Critical Data Studies (e.g. Kitchin, 2014; Leonelli, 2014; Rieder and Simon, 2016). While these fields have been instrumental in elaborating the neglected and problematic dimensions of Big Data, it remains an open question how and to what extent such insights become embedded in other fields, such as healthcare.

This paper critically reviews the epistemological claims and envisioned implications that accompany Big Data in the healthcare domain. The healthcare field is characterized by a strongly institutionalized set of epistemological principles and generally accepted scientific methodologies (Timmermans and Berg, 2003). Big Data challenge these principles and methodologies with the consequence that the epistemological implications of Big Data practices could be particularly profound. What we value as evidence and knowledge has implications for the way medical decisions are taken and healthcare is organized. Opening up the assumptions allows us to evaluate the role of Big Data in healthcare critically and open up opportunities for debate and fruitful intervention.

We base the paper on a systematic and comprehensive review of scientific editorials as these, in particular, summarize and reflect upon developments in the field. We focus on discourses surrounding Big Data in the analysis and construct five ideal-typical discourses based on a detailed analysis of the language conveyed in the editorials. The discourses show the diverse ways in which Big Data and the epistemological claims are conceptualized. We chose this focus as language is the medium through which people come to understand Big Data and it influences the way Big Data initiatives are performed and legitimated. Three questions guide our analysis:

(1) What Big Data discourses can be identified in scientific healthcare literature?
(2) How do the discourses conceptualize the meaning of evidence?
(3) What are the consequences of these conceptualizations for the way Big Data is understood in healthcare?

Big Data as material practice and semantic reality

Many authors have discussed the ambiguity surrounding the term Big Data. The term is often characterized by its volume, velocity and variety ('the 3Vs'; Mayer-Schönberger and Cukier, 2014). However, many believe that these three characteristics do not sufficiently capture Big Data. The 3Vs are thus often extended with extra 'V's, such as value, viability, variability, visualization and veracity (DeVan, 2016; Kitchin and McArdle, 2016). Others use different qualifications to characterize Big Data, such as exhaustively, relationality, extensionality and scalability (Boyd and Crawford, 2012; Kitchin and McArdle, 2016; Mayer-Schönberger and Cukier, 2014). Despite the many attempts, there is still no consensus about the term Big Data.

Inspired by the approach of Beer (2016) and Rudinow Saetnan et al. (2018), we conceptualize Big Data as a set of practices and ideas that exist in both (1) real material practice and in (2) a semantic reality. First, Big Data exist in specific actions, technologies and initiatives that are introduced to restructure healthcare. It is linked to the collection and aggregation of available data and correlation, pattern-recognition and predictive analyses. These data and analytics are subsequently used in real initiatives that aim to collect data, track, profile and predict behaviour, preferences and characteristics (Mittelstadt and Floridi, 2016). Second, Big Data exist in a semantic reality as it is something that we talk and write about in order to anticipate the (possible) effects. In this semantic reality, we envision and give meaning to the present and future of Big Data. Of course, the way we describe Big Data subsequently influences the way Big Data are performed and legitimated and vice versa.

In this paper and our analysis, we focus on the semantic reality of Big Data and discourses and metaphors. This is not to argue that detailed empirical investigations into material practices are less important. However, if we want to explore the implications of Big Data we also need a better understanding of how Big Data are discursively constructed. The crucial role of metaphors[3] in people's experience and sense-making of the world has been long recognized (Lakoff and Johnson, 2011) as metaphors play a large role in framing debates in particular ways. Metaphors are not neutral; they embody assumptions, imagined implications and impose opportunities and limitations (Puschmann and Burgess, 2014; Zinken et al., 2008). This makes metaphors especially valuable as we want to open up the epistemological claims and assumptions that accompany Big Data in healthcare.

Methodology

We conducted a comprehensive and systematic search of scientific literature to show the different ways in which Big Data and its epistemological claims are being articulated in the healthcare field. We chose this approach, because we did not want to miss major views and also gain insight in the relative spread of the articulations. Although our search of the literature fits the methodological approach of a systematic literature review, we subsequently departed from this approach in the interpretation and analysis of the results. While a 'traditional' review counts and synthesizes the results and provides an exhaustive *summary* of current evidence, we chose to follow a discourse analytic approach for the analysis because we wanted to move beyond a summary of results to provide an *interpretation* of the material (Dixon-Wood et al., 2006). The main advantage of this approach is that it combines the strengths of a systematic, thorough literature search with the explanatory power of interpretive analyses that provides new insights into a phenomenon.

Identifying relevant studies

A search term was composed with the help of a librarian to select the relevant studies. The search term covered terms related to (1) 'healthcare' and (2) 'Big Data' and related techniques, such as data mining. We wanted to be as inclusive as possible. The librarian and the first author looked for mentioning of the term Big Data in relevant studies and included those. Also, they started with a small list of techniques related to Big Data and iteratively added additional techniques to the search term if they were frequently mentioned in the found studies and resulted in relevant studies. The minimum requirement for inclusion was the mentioning of unusually large data sets or combinations of diverse types of data sets. We choose not to include the search term 'artificial intelligence' as this resulted in thousands of studies more for inclusion. In addition, we decided not to include 'knowledge', 'evidence' and related terms in the search profile, because we assumed that even studies that do not mention these terms can still make epistemological claims. The exact search terms are listed in Appendix 1. Eventually, we conducted the extensive search in Embase, Medline Ovid, Web of Science, Scopus, LISTA EBSCOhost and Google Scholar in January 2017.

We chose to limit our search to editorials from scientific journals in the healthcare domain because of their distinct characteristics. Editorials are expressions, reflections or commentaries on developments. They are a medium for editors, researchers and clinicians to communicate with peers and informed publics, as well as a forum for the explicit expression of beliefs and opinions (Loke and Derry, 2003; Miller et al., 2006). They can contain substantial scientific content, compelling messages, calls for action and discuss little known scientific facts with far-reaching consequences (Rousseau, 2009). They are usually written by the journals' editors or leading authors of the field. Editorials are often accessed and appear in well-regarded academic journals (Loke and Derry, 2003; Youtie et al., 2016). We selected editorials instead of viewpoints and opinion articles because we assume that editorials have a more critical role in defining the standpoint of the journal as compared to presenting the opinions of individuals. Lastly, editorials set the agenda for specific research fields and are a basis for future action. Hence, we believe that editorials capture Big Data discourses in the scientific community and have an important function in disseminating assumptions about Big Data in the healthcare domain.

Given the size of the original body of selected documents, further selection criteria were needed to obtain a manageable data set for detailed analysis. Hence, we chose to define a timeframe (2012–2016) for the review. As other studies have, we noticed an exponential increase in the number of publications about Big Data in general in 2012 (Youtie et al., 2016). Therefore, we choose 2012 as the starting point. Also, we included only English language editorials for practical reasons. If we could not find the editorial text online, we contacted the first author to gain access. In 24 instances, this did not work, and these documents were excluded because we could not access the full text.

The final selection of documents contained 1204 original documents. The first author of this paper read the title and abstract or the first and last paragraphs (if an abstract was unavailable) and excluded the irrelevant texts. Documents were excluded in close cooperation with the second and third authors because they either did not qualify as editorials or were outside the scope of this review (i.e. documents that were not about Big Data or were unrelated to health or healthcare). After screening, 206 editorials were eventually included for detailed review (see also Figure 1). Appendix 2 provides an overview of the included editorials.

Data analysis

The analysis was conducted in three phases. First, the first author randomly selected 20 editorials and flagged sections of interest. The authors of the paper discussed trends in the editorials and composed a list of questions

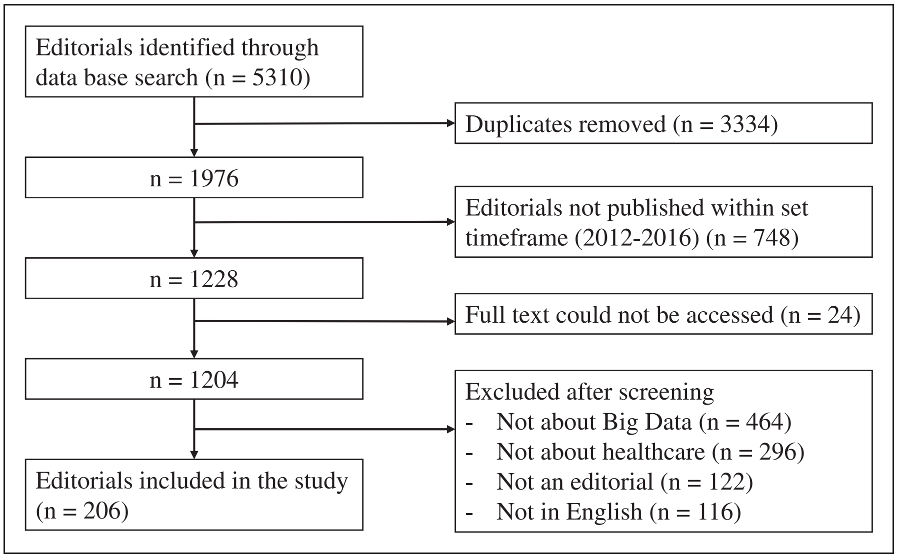


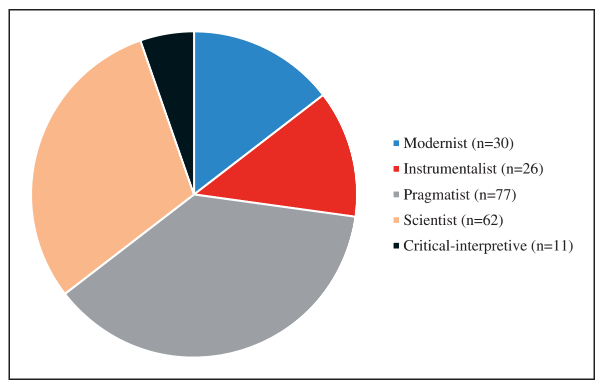
Figure 1. Selection of the editorials.

that would be relevant to answer for each editorial. Subsequently, the first and second author both analysed another 20 editorials and the list of questions was finalized. The list contained questions about (1) conceptualization of Big Data (e.g. how is Big Data described?), (2) the epistemological position (e.g. what is described as a good way of obtaining evidence/knowledge?), (3) the envisioned consequences (e.g. how are outcomes of Big Data used?) and (4) noticeable discursive elements, such as metaphors and surprising examples or comparisons. In the second phase, all remaining editorials were analysed with the finalized analytical scheme by the first author, second author and a junior researcher. The questions were answered for all the editorials and organized in a spreadsheet. Ten per cent of the editorials were also analysed by another member of the research team to ensure analytical consistency. Third, to organize and interpret the spreadsheet and to construct the ideal-typical discourses, the authors of this paper jointly tested, critically interrogated and experimented with the analytical themes and organization of results until consensus was reached about the structure and characteristics of the several discourses. This process eventually resulted in the construction of the five discourses.

Results

Description of data set and overview of findings

Based on our analysis, we were able to construct five ideal-typical discourses: modernist, instrumentalist, pragmatist, scientist and critical-interpretive. We drew inspiration for the names of the discourses from the relations



Graph 1. Presence of the ideal-typical discourses in the editorials.

we saw between implicit assumptions about evidence and knowledge and diverse philosophical and epistemological positions. The discourses were distributed over the editorials in the following way: modernist ($n = 30$), instrumentalist ($n = 26$), pragmatist ($n = 77$), scientist ($n = 62$) and critical-interpretive ($n = 11$; see Graph 1). These discourses should be viewed as ideal-types, meaning that some editorials consist of combinations of various discourses. Co-occurrence especially consisted between the instrumentalist and pragmatist discourses ($n = 16$) and between the modernist and pragmatist discourses ($n = 12$). The modernist and critical-interpretive discourses and the instrumentalist and critical-interpretive discourses did never co-occur in one editorial.

We summarized the discourses and their main characteristics in Table 1. We will describe the five ideal-typical discourses in more detail below. In our description of the discourses, we will highlight one metaphor

Table 1. Overview of the discourses.

Ideal-typical discourse	Modernist	Instrumentalist	Pragmatist	Scientist	Critical-interpretive
Conceptualizing Big Data					
Big Data are described as	Large amounts of data that can be analysed	Analytic techniques	A useful (managerial) instrument for decision-making	A trend that deals with data collection, analysis and outcomes more flexibly	A trend that oversimplifies reality
Evaluation of Big Data	Positive	Positive	Positive	Critical	Critical
Recommendations for further development	Start to use Big Data in healthcare	Enhance and develop the Big Data techniques	Implement Big Data in healthcare	Be (extremely) careful with the use of Big Data	Discuss the negative consequences of Big Data
Non-use of Big Data is explained in terms of	Not discussed	Techniques do not work sufficiently	Implementation of problems	Lack of performance (as traditional studies perform better)	Negative consequences for individuals and society
Epistemological position					
Inference from data	Direct, data equal knowledge	Direct, data equal knowledge (that we can see through the techniques)	Direct, data equal knowledge (if useful in practice)	Indirect, data interpretation involves scientific methodology (hypothesis testing)	Indirect, data interpretation involves critical thinking
Epistemological claim	Big Data offers reliable information	Big Data offers increasingly more reliable information as the techniques improve	Big Data can offer reliable information in some situations	Big Data can be useful if strict criteria are met	Big Data will always generate limited evidence
Presumed reliability of Big Data	High	High	High	Medium–Low	Low
Summarizing metaphor	Capturing data	Illuminating data	Harnessing data	Selecting data	Constructing data
Consequences					
Presumed consequences of Big Data	Revolutionary amount of new knowledge	New predictions and increased understanding to solve persistent problems	Improved problem-solving and decision-making in healthcare	Inconclusive and misguided information, if Big Data are not properly used	Inconclusive and misguided information and unfair outcomes

that is particularly apt to illustrate the epistemological positions of each specific discourse.

The modernist discourse: Capturing data

The conceptualization of Big Data. In this ideal-type, Big Data are often not defined, but the editorials link it to large amounts of data. Big Data are described as a positive development and the editorials stress the beneficial effects of Big Data. They state, for example, that it will lead to proactive, predictive, preventive, participatory and patient-centred health (Shah and Tenenbaum, 2012; Weinstein, 2016). However, the precise meaning of these statements often remains unclear and ambiguous, as they are not discussed further.

The editorials unanimously and unambiguously recommend the use of Big Data in healthcare. This is emphasized by three rhetorical techniques. First, the tone of these editorials is optimistic, signified by such words as 'explosion', 'revolutionizing', and 'world-changing possibilities'. Big Data are presented as innovative and as a rupture with the past that will radically transform healthcare (Restifo, 2013; Weinstein, 2016). Secondly, a sense of urgency is created in the editorials as they often draw a contrast between the medical domain and other sectors that supposedly already take advantage of Big Data. The medical domain is presented as slow, conservative and old-fashioned, while other domains are already taking Big Data analytics for granted. This discursively constructs the field of medicine and its current approaches as unsustainable and outdated (MacRae, 2012; Risoud et al., 2016). Third, there is almost no attention for the negative sides of Big Data, such as potential issues with privacy, consequences of shifting power-relations or for practical questions concerning implementation. Illustrative of this position is the almost complete lack of non-use of Big Data as a theme in this discourse.

Epistemological assumptions. Capturing data is the metaphor (Figure 2) that most clearly illustrates the epistemological assumptions in the modernist discourse. First, because the modernist discourse assumes data to exist in the world and to have inherent value (like a butterfly or other natural resources). The assumptions are that the data can be captured and that this results in new insights, evidence and practices. Second, the metaphor aptly illustrates the epistemological assumptions in this discourse because capturing is a relatively simple act that also leaves the data itself unaffected, which shows the ease in which Big Data are portrayed in these editorials to be able to arrive at knowledge. This process is viewed in such simplistic terms that data seem to equal knowledge. This creates the idea that only 'capturing data' already leads to new knowledge.

Figure 2. Capturing data metaphor.

Consequences. The modernist discourse strives for a radical change as the traditional ways of knowledge production in the medical domain are rejected. Editorials in the modernist discourse aim to overthrow the status quo in order to transform knowledge production in healthcare radically. Big Data are seen as a legitimate source of knowledge in these editorials because Big Data are argued to lead to more timely and reliable knowledge that is viewed as immediately useful in practice. However, the discourse seems to be naïve in the sense that it only addresses grand visions and is not concerned with, for example, the practical development and application of Big Data, nor with the societal effects.

The instrumentalist discourse: Illuminating data

The conceptualization of Big Data. In this ideal-type, Big Data are understood in terms of a range of analytical techniques, such as pattern-recognition, data mining and machine learning (Amato et al., 2013). The editorials have a positive tone and describe ways in which these Big Data techniques can aid healthcare, for example by predicting disease outcomes and increasing the understanding of the causes of diseases (Belgrave et al., 2014; Van De Ville and Lee, 2012). The editorials typically discuss how analytic techniques should be used and how they can be improved. The editorials contain advice on how one should deal with the missing data, correlated features and replication and separation of training and validation sets.

The editorials recommend that Big Data techniques should be developed and enhanced to gain better results. Editorials in this discourse place a high value on experimentation. For example, innovative studies in which Big Data techniques are used for brain decoding and the development of clinical decision support systems are presented (Najarian et al., 2013; Van De Ville and Lee, 2012). Using Big Data techniques for these purposes is by no means standard practice, but by trying out and experimenting with data analytic processes, the techniques are improved. Illustratively, terms like improving, experimenting, exploring, developing and learning frequently occur in the instrumentalist editorials.

Epistemological assumptions. The illuminating data metaphor (Figure 3) best represents the epistemological assumptions in the instrumentalist discourse and is exemplified by phrases such as 'casting light' and 'highlighting' in the editorials. Similar to the modernist discourse, in the instrumentalist discourse data seem to exist in the world and are viewed as having an intrinsic value. However, the process of knowledge discovery through Big Data is depicted in less simplistic terms than in the modernist discourse, as the editorials emphasize that information can only be extracted from highlighting the data with specific analytic techniques so that patterns in the data can be seen (Amato et al., 2013; Rosenstein et al., 2014). This is an indirect critique of the more traditional methods for knowledge generation, which are implicitly depicted as outdated and inefficient. The editorials thus suggest that by constructing and positioning the 'light sources' (e.g. the analytic techniques), we are increasingly able to 'see' the data and emerging trends within them. This means that knowledge improves together with the set of analytical techniques.

Consequences. The instrumentalist discourse promotes the use of Big Data techniques in healthcare as they become a reliable source for decision-making. Less radically than the modernist discourse, editorials in this discourse still argue for a change of the ways knowledge is obtained in healthcare, as Big Data are expected to solve persistent problems in healthcare. The discourse seems to envision Big Data as a tool to solve problems and the tool is valid to the extent that it helps to make accurate predictions and increases our understanding. However, similar to the modernist discourse, the instrumentalist discourse also neglects the broader implications and potential societal effects of the use of Big Data techniques.

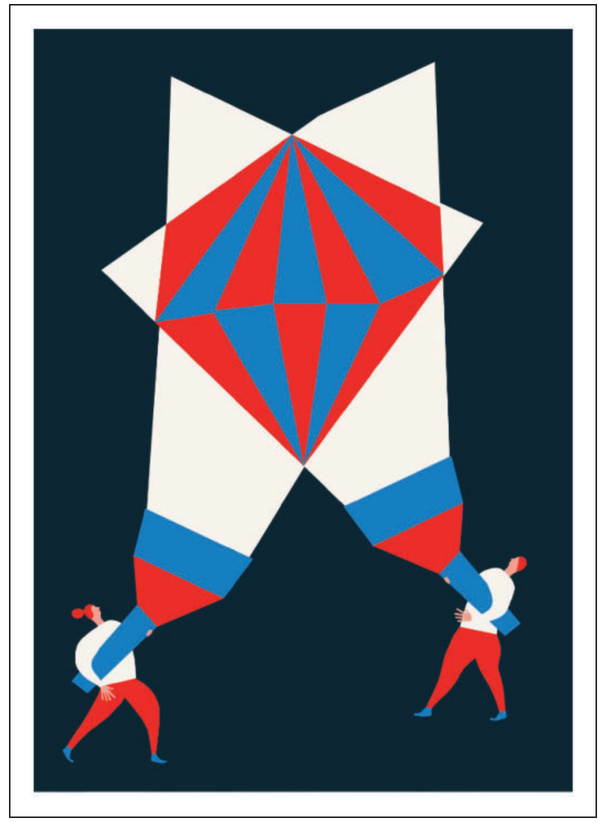


Figure 3. Illuminating data metaphor.

The pragmatist discourse: Harnessing data

The conceptualization of Big Data. In this ideal-type, Big Data are conceptualized as a useful (managerial) instrument for problem-solving and decision-making in healthcare (Garrison, 2013; Klonoff, 2013; Potters et al., 2016). Big Data are discursively constructed in the editorials as a phenomenon that is already here and is likely to stay (Basak et al., 2015; Ghani et al., 2014; Hay et al., 2013). Big Data are described as a positive development. However, in this discourse, people are presumed to have a significant influence on the way Big Data take shape, as opposed to the more technological determinist pattern of thinking that characterizes the modernist discourse.

The editorials in this discourse primarily focus on how Big Data should be implemented and describe the steps for successful implementation. They discuss, for example, the training, recruitment and the introduction of data scientists or knowledge engineers, cultural factors that need to change in healthcare, new rules and regulations that have to be made, the adoption of new platforms and information systems, and how access should be gained to the data and analytics (Cases et al., 2013; Kottyan et al., 2015; Narula, 2013; Potters et al., 2016). The editorials do mention concerns and other challenges that need to be overcome or solved, as the following quote from McNutt et al. (2016: 914) illustrates:

> 'We envision future systems that incorporate [Big Data] decision support models into the clinical systems in ways that enable clinicians to improve both the quality and the safety of care they give and the efficiency with which they give it. To reach this vision, there remain technological needs and human challenges to overcome.'

Epistemological assumptions. The metaphor of 'harnessing data' (Figure 4) best illustrates the ideas and assumptions about Big Data in the pragmatist discourse. Similar to the previous discourses, data continue to be described as something 'out there', simply existing in the world. The data are viewed as valuable as they can be translated into information and knowledge. Different is that this discourse sees traditional scientific and Big Data methods as complementary approaches that can both generate 'evidence' and have practical relevance (Basak et al., 2015; Klonoff, 2013). A more pragmatic attitude towards evidence seems dominant as evidence is not strictly related to scientific processes. There are no fundamental objections against using Big Data outcomes. Big Data are viewed as beneficial whenever it helps to gain knowledge about situations that traditional scientific methods cannot study and decision-makers pragmatically make choices on the basis of the available evidence. Discussions about the status of the outcomes of traditional scientific studies and Big Data analyses disappear to the background in this discourse, as the actionable character is emphasized.

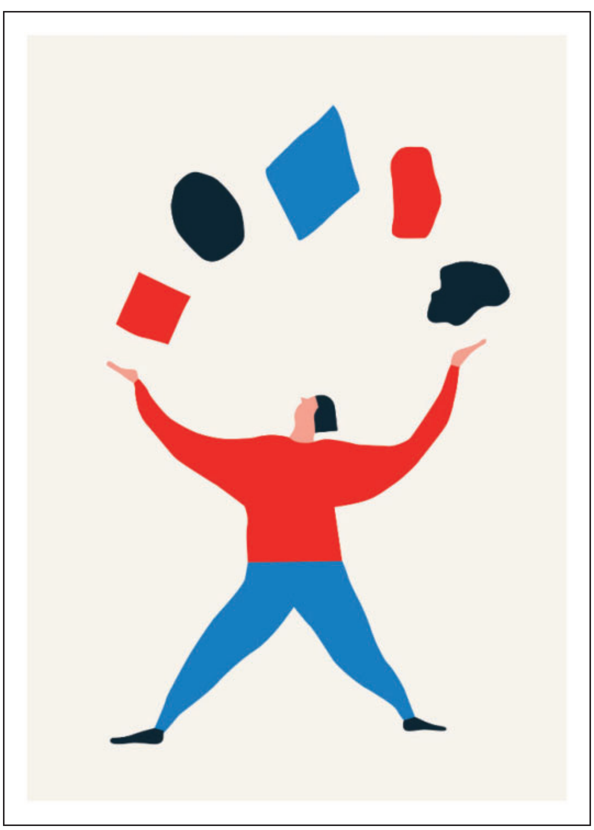


Figure 4. Harnessing data metaphor.

Consequences. Similar to the instrumentalist discourse, the pragmatist discourse envisions a change in the way decisions are taken as Big Data offer more knowledge than currently is available and can generate useful new insights for healthcare practice. Big Data are seen as a valuable source for decision-making next to traditional knowledge producing approaches. This discourse deals – more than the previous discourses – with some of the practical issues surrounding Big Data implementation (such as the recruitment of data scientists). However, the epistemological and normative changes that Big Data bring are not addressed.

The scientist discourse: Selecting data

The conceptualization of Big Data. In this ideal-type, Big Data are described as a new trend that deals with data collection, analysis and outcomes in a less rigorous way than scientific methodologies do. The editorials mention that Big Data can be useful in some situations because of its potential to identify valuable research directions, for hypothesis-generation and exploration of massive data sets (Khoury and Ioannidis, 2014; Krakoff and Phillips, 2016). It can thus only be used as exploratory, hinting at possible directions for traditional research designs. The tone of the editorials is critical, especially compared with the modernist discourse, and Big Data are seen as a potentially dangerous development.

The editorials argue for caution with regards to Big Data and claim that traditional scientific methods will remain essential despite the arrival of Big Data methodologies. The editorials try to distinguish 'proper' from erroneous science. They do this, for example, by comparing Big Data outcomes and findings from RCTs (Freeman and Saxon, 2015). Some editorials mention the limitations of traditional studies. For example, they state that RCTs are costly or not always possible because of ethical considerations (Freeman and Saxon, 2015; Leem, 2016). However, the consensus seems to be that despite the potential of Big Data as a starting point for research, it always needs to be followed by more substantive research. Or as Khoury and Ioannidis (2014: 1054) state in their editorial: 'We should embrace (and not run away from) principles of evidence-based medicine.'

Epistemological assumptions. The epistemological assumptions about Big Data within this discourse can be summarized by the metaphor of 'selecting data' (Figure 5). The notion that Big Data can lead to reliable and valid knowledge is questioned and sometimes outright denied in the editorials. Two arguments are frequently made. First, the editorials stress that data are essential to arrive at knowledge. However, data are not viewed as pre-existing in the world. As such, they cannot simply be captured, illuminated or harnessed, but need to be selected and processed via specific methods. This position is reinforced by statements like 'garbage in, garbage out' (denoting the idea that the lack of selecting 'high-quality' data from the masses of available, often poor quality data leads to useless analyses), or by presenting the data of Big Data as erroneous or as a 'dumping site' (Brown, 2016; Patrick, 2016). Through discursively oppositioning high-quality data with 'garbage', the editorials point to the need to have the proper or right procedures for data gathering and analysis in place. Such procedures are meticulous and less easily abandoned than presumed in, for example, the modernist discourse. Second, the editorials problematize the assumption that more data equal better knowledge. This idea is widespread in the modernist, instrumentalist and – to some extent – pragmatist discourses. According to editorials in the scientist discourse, this assumption is wrong. As Onukwugha (2016: 92) explains:

> 'We cannot assume that more data necessarily means more information. Indeed, as the volume of data increases, it will be important to pay continued (or more) attention to established concerns regarding measurement, bias, and fallacies relevant to empirical analysis and interpretation.'

Despite the criticism, the epistemological position is similar to the modernist and instrumentalist discourses as the positivistic notion that truth can be found in data is also present. However, in the modernist and – to some extent – instrumentalist discourse there seem to be an acceptance of a rather naïve empiricism that, according to the scientist discourse is too simplistic. The scientist discourse argues that, for example, Big Data can be informative, but never capture a whole domain and that there remains a need for hypotheses and theory. So, evidence is assumed to be developed only by correctly applying the scientific method. Just experimenting with Big Data can lead to wrong conclusions (Gomella, 2016).

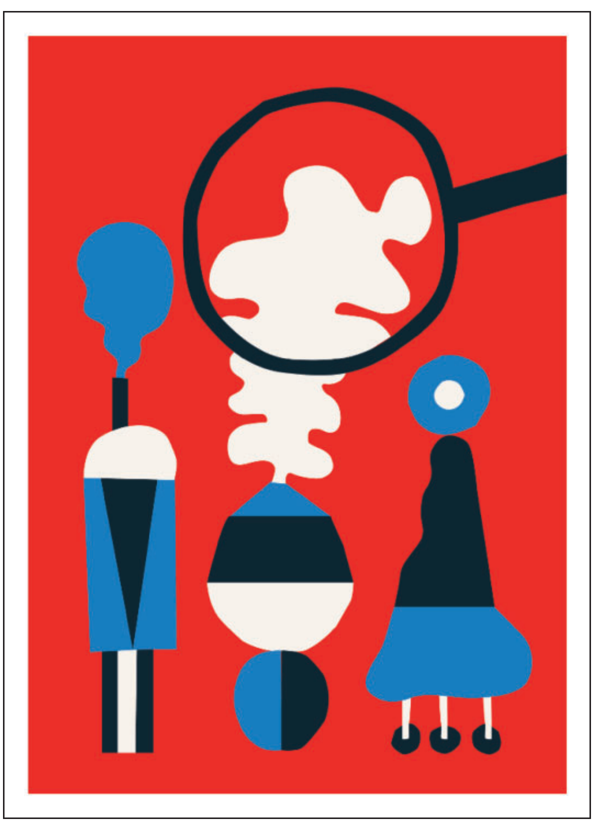


Figure 5. Selecting data metaphor.

Consequences. The scientist discourse argues against a radical change in healthcare as according to this discourse, Big Data are not a reliable source of knowledge. The only proper knowledge seems to be scientific knowledge and such knowledge can only come from the use of strict scientific methods. The consequences of Big Data would be erroneous evidence and knowledge with possibly large, detrimental effects. This discourse discusses in-depth the epistemological concerns and how Big Data related to traditional structures for knowledge generation.

The critical-interpretive discourse: Constructing data

The conceptualization of Big Data. In this ideal-type, Big Data and data are presented as an oversimplified presentation of reality. The critical-interpretive discourse incorporates diverse forms of criticisms. Generally, the editorials share a concerned tone and their criticisms are both epistemological and societal.

The editorials advocate discussion on the position of Big Data in our society as a whole. Two lines of critique can be distinguished in this discourse. First, the simplicity of data is frequently addressed. Big Data are dismissed because it is a reductionist and oversimplified presentation of reality, unable to adequately capture and account for the richness and diversity of human experience. Editorials make this point by describing data that are missing in Big Data sets and by stressing the importance of personal experience, objectives and preferences (Pope et al., 2014; von Gunten et al., 2016; Zurlinden, 2016). Second, the editorials stress the normative aspects of Big Data and point out that these aspects are often overlooked or neglected. The editorials, for example, focus on the danger of Big Data that is not being interpreted by physicians and warn that Big Data can be a first step for 'dangerous' automatic decision models. As Von Gunten et al. (2016: 1240) state: 'It [Big Data outcomes] must be interpreted by a seasoned clinician with critical thinking skills.'

Epistemological assumptions. The epistemological assumptions that characterize editorials in this discourse can be best understood via the metaphor of 'constructing data' (Figure 6). In terms of epistemological assumptions, the critical-interpretive discourse is most distinctive from the other discourses as it reasons from a different set of epistemological assumptions (building on constructivist traditions in philosophy of science as opposed to positivist approaches). Consequentially, data are no longer presented as something given that can be captured or illuminated, but understood as the result of the social and political processes that created them. As Pope et al. (2014: 68) state: 'We must remember that all data – big or small – are socially constructed.' This perspective means a recognition that data always emphasize certain aspects of the world while leaving out other elements. Importantly, the constructed data present an image, but editorials in this discourse warn that this image can never be complete. This discourse can especially be contrasted with the modernist discourse, in which the ideal of 'complete knowledge' is maintained. Big Data, therefore, according to the critical-interpretive discourse, will always generate limited knowledge and data have to be handled with care.

Consequences. The critical-interpretive discourse warns for the limitations of Big Data. According to this discourse, while Big Data create new possibilities for generating knowledge, the use of these possibilities is not seen as a positive change. The starting point is that it is better not to use Big Data (or at most only with great restraint). The consequences of Big Data would be that limited data are extrapolated and would lead to erroneous outcomes that could cause harm to people and healthcare systems. In addition, if people are not able to recognize the fact that data are constructed, for example, by the use of automated decision models, essential aspects of care would be lost.

Figure 6. Constructing data metaphor.

Discussion

Reviewing literature is a first step in gaining a better understanding of the epistemological implications of Big Data in healthcare. Based on a systematic literature search and consecutive interpretive analysis, we constructed five ideal-typical discourses of Big Data in healthcare. These five discourses all highlight particular aspects of Big Data, neglecting others, and thereby frame Big Data and its (epistemological) implications in specific ways. This study is vital because discourses and metaphors pre-structure the way that the material practices of Big Data take shape. As such, they are highly consequential in shaping current and future debates on Big Data. In this discussion, we will take the next step by drawing attention to the political dynamics of the discourses.

We build on insights from STS and Critical Data Studies to point to issues that have been ignored or neglected in the current construction of the Big Data debate in healthcare editorials. We end with suggestions for future research.

We noticed that the discourses that frame Big Data in positive terms (modernistic, instrumentalist and pragmatist) were more present in our empirical material ($n = 133$, 64.6%). These discourses seem to reinforce each other in the idea that Big Data result in valid knowledge and that massive data sets and predictive analytics reflect the truth. These grand promises could explain the strong positive rhetoric that surrounds the term Big Data and that drives implementation of Big Data initiatives in healthcare. The corresponding metaphors of capturing, illuminating and harnessing data all embody closely related epistemological expectations. Data are presented as benign, objective, an asset for organizations, and not something that should be questioned. Big Data are seen to settle previously unsolvable problems. The three discourses all view the advancement of Big Data into healthcare as inevitable (Mayer-Schönberger and Cukier, 2014; Murdoch and Detsky, 2013), with the instrumentalist discourse more concerned about the development of the analytic techniques and the pragmatist discourse more concerned about the implementation of Big Data.

The discourses that frame Big Data in more critical terms (scientist and critical-interpretive) were less present in the editorials ($n = 73$, 35.4%). They both challenge the objectivity, effectivity and serviceability claims that are dominant in the positive discourse, do not view Big Data as inevitable and pose alternative possibilities. This is important for healthcare, as they make sure we reflect on Big Data knowledge. However, both discourses do this from different implicit philosophical positions (positivist and constructivist). Their metaphors of selecting and constructing data illustrate another political message that frame Big Data as limited, and claims that positive Big Data discourses obscure the often serious implications for expertise and evidence.

Especially editorials in the critical-interpretive discourse were limited ($n = 11$, 5.3%). This is an interesting observation in the light of the increased attention for the problematic assumptions and epistemological difficulties of Big Data in fields such as STS and Critical Data Studies, often offering fundamental criticisms about the claims and expectations surrounding Big Data. For example, that although data may appear objective, they are still constructed through subject–technology interactions (Boyd and Crawford, 2012; Dalton and Thatcher, 2014; Kitchin and Lauriault, 2014). An important conclusion that can be drawn from our analysis is that such work has not broadly infiltrated the domain of medical editorials.

We argue that the healthcare field would benefit from a more prominent critical-interpretive discourse, as three important issues would be neglected (as they are not addressed by the other discourses): (1) the normative assessment of Big Data, for example, the role that automatic decision models should play in the doctors' office and issues related to data access and consent (Mittelstadt and Floridi, 2016). (2) Reflection on the situatedness of data. Data do not speak for themselves and we must remember that they are always an oversimplification of reality. Reflection on what particular aspects of a phenomenon are emphasized in the data and what aspects are occluded is therefore crucial (Boyd and Crawford, 2012; Mittelstadt and Floridi, 2016). (3) The social and political processes that create Big Data. While Big Data and data may seem objective to many, they still are subjective and contain biases and other limitations which should be opened up (Boyd and Crawford, 2012). We believe that the pragmatist discourse deals with the first issues too pragmatically and the scientist discourse with the last issues too statically and without enough attention for the social dynamics. Subsequently, the healthcare field would benefit from more critical reflection and intervention.

Based on this review, we stress that the epistemological discussion in healthcare needs to be developed further and that we have to find ways to better integrate aspects of the critical-interpretive discourse in the healthcare domain. Based on this paper, we suggest the following directions for further research:

1. Further study into the five ideal-typical discourses could provide important insights into the ways (and extent in which) similar discourses and dynamics are also noticeable in other disciplines. Quantitative approaches could investigate correlations between the background of editors/authors and the discourses they endorse.
2. As discourses are not only part of editorials, but also of broader cultural discussions, future research could study the various ways in which the semantic realities of Big Data intersect with material practices and vice versa. Especially warranted are comparative studies that open up the ways Big Data are depicted in different cultural domains and the sociotechnical imaginaries (Jasanoff and Kim, 2015) in which these depictions are embedded.
3. Empirical reflections on the material practices of Big Data are warranted as well. Discourses and sociotechnical imaginaries are still part of theoretical

discussions, while at the same time many Big Data initiatives are started in healthcare. Studying such initiatives ethnographically is likely to provide highly valuable insights into the dynamic encounters between data and healthcare.

Conclusion

The fields of STS and Critical Data Studies have been instrumental in opening up discussions about the epistemological and ethical implications of an emerging field of practices, captured under the umbrella term 'Big Data'. On the basis of this study, we have to conclude that these reflections have not been embedded in the healthcare field in any substantial way. Based on a systematic analysis of scientific editorials, we constructed five ideal-typical discourses to gain a better understanding of how Big Data are discursively constructed. We observed that editorials in the critical-interpretive discourse were limited (only 5.3%). We conclude that the healthcare field would benefit from a more prominent critical-interpretive discourse, since important reflections on the normativity and situatedness of Big Data, as well as the social and political processes that create Big Data, are not addressed by the other discourses.

Acknowledgements

We thank the anonymous reviewers for their valuable feedback on earlier versions of the paper. Furthermore, we would like to thank Sue Doeksen for the beautiful illustrations, the librarian Wigor Bramer for the help with the search term and Mark Zijdemans for his help in analyzing the editorials and the members of the Healthcare Governance group at the Erasmus University Rotterdam for their valuable comments and the discussions.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Notes

1. We see Big Data as a set of practices and technologies that is discursively framed under the umbrella term 'Big Data'. We do not see Big Data as a coherent unity and therefore set Big Data in plural form.
2. We recognize that terms like 'data', 'information', 'knowledge' and 'evidence' are notoriously ambiguous as many definitions circulate. The terms are also used in different ways in the various discourses that we outline in this paper. In principle, we use the terms 'data', 'information' and 'knowledge' hierarchically. Data (points) become information after they are grouped and eventually knowledge when they are further contextualized. The term 'evidence' originates from a different tradition and is therefore primarily used to refer to discussions about evidence-based medicine. In our description of the discourses, we follow the authors' use of the terms.
3. Two recent studies explored metaphors used to describe Big Data in popular mass media and business press. The first study by Puschman and Burgess (2014) recognizes two Big Data metaphors in mass media. Both dominant metaphors stress the idea that data accurately reflect nature, society and culture, and that the presented units (e.g., data) are comparable and the results are reproduced. The other study (Maiers, 2017) examined business press and noticed the frequent use of oriental metaphors. Maiers recognized a vertical direction in the metaphors (e.g. deep analytics, data mining, and drilling down) that suggest the assumption that by going deeper, more details, accuracy and precision can be found. We were surprised by the strength of the positivistic ideas related to these metaphors of Big Data because these are not only part of popular mass media and the business press, but are also actively embraced by many medical researchers and are recognizable in the editorials of renowned scientific journals.

References

Amato F, López A, Peña-Méndez EM, et al. (2013) Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine* 11(2): 47–58.

Anderson C (2008) The end of theory: The data deluge makes the scientific method obsolete. In: *Wired*. Available at: www.wired.com/2008/06/pb-theory/ (accessed 5 September 2017).

Basak SC, Vracko M and Bhattacharjee AK (2015) Big data and new drug discovery: Tackling 'Big Data' for virtual screening of large compound databases. *Current Computer-Aided Drug Design* 11(3): 197–201.

Beer D (2016) How should we do the history of Big Data? *Big Data & Society* 3(1): 1–10.

Belgrave D, Simpson A and Custovic A (2014) Challenges in interpreting wheeze phenotypes: The clinical implications of statistical learning techniques. *American Journal of Respiratory and Critical Care Medicine* 189(2): 121–123.

Berwick DM (2016) Era 3 for medicine and health care. *JAMA: The Journal of the American Medical Association* 315(13): 1329.

Boyd D and Crawford K (2012) Critical questions for Big Data. *Information, Communication & Society* 15(5): 662–679.

Brown ML (2016) Can't you just pull the data? The limitations of using of the electronic medical record for research. *Pediatric Anesthesia* 26(11): 1034–1035.

Cases M, Furlong LI, Albanell J, et al. (2013) Improving data and knowledge management to better integrate health care and research. *Journal of Internal Medicine* 274(4): 321–328.

Crawford K, Miltner K and Gray ML (2014) Critiquing Big Data: Politics, ethics, epistemology. *International Journal of Communication* 8: 1663–1672.

Dalton C and Thatcher J (2014) What does a critical data studies look like, and why do we care? Seven points for a critical approach to 'Big Data'. In: *Space and Society Open Site*. Available at: http://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher/ (accessed 6 April 2017).

DeVan A (2016) The 7 V's of Big Data. In: *Impact Radius*. Available at: www.impactradius.com/blog/7-vs-big-data/ (accessed 11 September 2017).

Dixon-Wood M, Cavers D, Agarwal S, et al. (2006) Conducting a crticial interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology* 6(35): 1–13.

Felder M and Meerding WM (2017) *Een toekomst voor evidence-based medicine? Achtergrondstudie bij het advies ''Zonder context geen bewijs. Over de illusie van evidence-based practice in de zorg.''*. The Hague: Raad voor Volksgezondheid en Samenleving.

Freeman JV and Saxon L (2015) Remote monitoring and outcomes in pacemaker and defibrillator patients. *Journal of the American College of Cardiology* 65(24): 2611–2613.

Garrison LP (2013) Universal health coverage – Big thinking versus Big Data. *Value in Health* 16(1): S1–S3.

Ghani KR, Zheng K, Wei JT, et al. (2014) Harnessing Big Data for health care and research: Are urologists ready? *European Urology* 66(6): 975–977.

Gomella LG (2016) Big Data equals big challenges for prostate cancer. *The Canadian Journal of Urology* 23(4): 8329.

Greenhalgh T, Howick J and Maskrey N (2014) Evidence based medicine: A movement in crisis? *BMJ (Clinical Research ed.)* 348: g3725.

Hay SI, George DB, Moyes CL, et al. (2013) Big Data opportunities for global infectious disease surveillance. *PLoS Medicine* 10(4): 1–4.

Jasanoff S and Kim S-H (eds) (2015) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. Chicago, IL: University of Chicago Press.

Khoury MJ and Ioannidis JPA (2014) Big Data meets public health. *Science* 346(6213): 1054–1055.

Kitchin R (2014) Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1): 1–12.

Kitchin R and Lauriault TP (2014) Towards critical data studies: Charting and unpacking data assemblages and their work. In: Eckert J, Shears A and Thatcher J (eds) *Geoweb and Big Data*. Lincoln, NE: University of Nebraska Press.

Kitchin R and McArdle G (2016) What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society* 3(1): 1–10.

Klonoff D (2013) Twelve modern digital technologies that are transforming decision making for diabetes and all areas of health care. *Journal of Diabetes Science and Technology* 7(2): 291–296.

Kottyan LC, Weirauch MT and Rothenberg ME (2015) Making it big in allergy. *Journal of Allergy and Clinical Immunology* 135(1): 43–45.

Krakoff LR and Phillips RA (2016) Blood pressure variability. *Journal of the American College of Cardiology* 68(13): 1387–1388.

Lakoff G and Johnson M (2011) *Metaphors We Live by: With a New Afterword*, 6th ed. Chicago, IL: University of Chicago Press.

Leem JG (2016) Big Data and pain. *The Korean Journal of Pain* 29(4): 215.

Leonelli S (2014) What difference does quantity make? On the epistemology of Big Data in biology. *Big Data & Society* 1(1): 205395171453439.

Loke YK and Derry S (2003) Does anybody read 'evidence-based' articles? *BMC Medical Research Methodology* 3(14): 1–6.

McNutt TR, Moore KL and Quon H (2016) Needs and challenges for Big Data in radiation oncology. *International Journal of Radiation Oncology*Biology*Physics* 95(3): 909–915.

MacRae CA (2012) Pattern recognition: Combining informatics and genetics to re-evaluate conduction disease. *Heart* 98(17): 1263–1264.

Maiers C (2017) Data talk: Metaphors of an epistemological landscape. *Annual meeting of the Society for Social Studies of Science*. Boston: USA 31 August 2017.

Mayer-Schönberger V and Cukier K (2014) *Big Data: A Revolution that will Transform How We Live, Work, and Think*. Boston, MA: Mariner Books, Houghton Mifflin Harcourt.

Miller FA, Ahern C, Smith CA, et al. (2006) Understanding the new human genetics: A review of scientific editorials. *Social Science & Medicine* 62(10): 2373–2385.

Mittelstadt BD and Floridi L (2016) The ethics of Big Data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics* 22(2): 303–341.

Murdoch TB and Detsky AS (2013) The inevitable application of big data to health care. *JAMA: The Journal of the American Medical Association* 309(13): 1351–1352.

Najarian K, Ward KR and Shirani S (2013) Biomedical signal and image processing for clinical decision support systems. *Computational and Mathematical Methods in Medicine* 1: 1–2.

Narula J (2013) Are we up to speed? *JACC: Cardiovascular Imaging* 6(11): 1222–1224.

Onukwugha E (2016) Big Data and its role in health economics and outcomes research: A collection of perspectives on data sources, measurement, and analysis. *Pharmacoeconomics* 34(2): 91–93.

Patrick K (2016) Harnessing Big Data For health. *Canadian Medical Association Journal* 188(8): 555.

Pope C, Halford S, Tinati R, et al. (2014) What's the big fuss about 'Big Data'? *Journal of Health Services Research & Policy* 19(2): 67–68.

Potters L, Ford E, Evans S, et al. (2016) A systems approach using big data to improve safety and quality in radiation oncology. *International Journal of Radiation Oncology*Biology*Physics* 95(3): 885–889.

Puschmann C and Burgess J (2014) Metaphors of Big Data. *International Journal of Communication* 8(1): 1690–1709.

Restifo NP (2013) A 'Big Data' view of the tumor 'Immunome'. *Immunity* 39(4): 631–632.

Rieder G and Simon J (2016) Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data. *Big Data & Society* 3(1): 1–6.

Risoud M, Bonne NX and Vincent C (2016) Big Data: Coming soon to ENT. *European Annals of Otorhinolaryngology, Head and Neck Diseases* 133(3): 157.

Rosenstein BS, West CM, Bentzen SM, et al. (2014) Radiogenomics: Radiobiology enters the era of Big Data and team science. *International Journal of Radiation Oncology*Biology*Physics* 89(4): 709–713.

Rousseau R (2009) The most influential editorials. *Celebrating Scholarly Communication Studies* 1(1): 47–53.

Rudinow Saetnan A, Schneider I and Green N (2018) The politics of Big Data: Principles, policies, practices. In: Rudinow Saetnan A, Schneider I and Green N (eds) *The Politics of Big Data: Big Data, Big Brother?*. Abingdon / New York, NY: Routledge.

Shah NH and Tenenbaum JD (2012) The coming age of data-driven medicine: Translational bioinformatics' next frontier. *Journal of the American Medical Informatics Association* 19(1): e2–e4.

Timmermans S and Berg M (2003) *The Gold Standard: The Challenge of Evidence-based Medicine and Standardization in Health Care*. Philadelphia, PA: Temple University Press.

Van De Ville D and Lee SW (2012) Brain decoding: Opportunities and challenges for pattern recognition. *Pattern Recognition* 45(6): 2033–2034.

Von Gunten CF, Teno JM and Morrison R (2016) Big Data and end-of-life care: Promise and peril. *Journal of Palliative Medicine* 19(12): 1240–1240.

Weinstein JN (2016) An 'Industrial Revolution' in health care: The data tell us the time has come. *Spine* 41(1): 1–2.

Youtie J, Porter AL and Huang Y (2016) Early social science research about Big Data. *Science and Public Policy* 44(1): 65–74.

Zinken J, Hellsten I and Nerlich B (2008) Discourse metaphors. In: Frank R, Dirven R, Ziemke T, et al. (eds) *Body, Language, and Mind. Vol. 2: Sociological Situatedness*. Berlin: Mouton.

Zurlinden J (2016) Health professions education: Make Big Data meaningful. *The Journal of Continuing Education in Nursing* 47(5): 199–200.

Appendix 1: Search terms

Embase.com

('machine learning'/de OR 'automated pattern recognition'/de OR 'automatic speech recognition'/de OR 'Bayesian learning'/de OR 'data mining'/de OR 'classification algorithm'/de OR 'computer heuristics'/de OR 'knowledge discovery'/de OR 'learning algorithm'/de OR 'network learning'/de OR 'online analytical processing'/de OR 'relevance vector machine'/de OR 'support vector machine'/de OR 'supervised machine learning'/de OR 'unsupervised machine learning'/de OR (('data analysis'/de OR 'data processing'/de OR 'pattern recognition'/de) AND ('correlation analysis'/de OR 'automation'/de OR bioinformatics/de OR 'medical technology'/de OR 'medical informatics'/de)) OR (((machine OR Bayesian OR network OR Autonom* OR semi-supervis* OR semisupervis* OR unsupervis*) NEAR/3 learning) OR data-mining OR text-mining OR ((automat* OR algorithm* OR bioinformat*) NEAR/6 (pattern* OR speech OR parameter*) NEAR/6 (recogni* OR select*)) OR ((classif* OR learning) NEAR/6 (algorithm* OR automat*)) OR (computer* NEAR/3 heuristic*) OR (knowledge* NEAR/3 discover*) OR (online NEAR/6 analytic* NEAR/6 process*) OR (vector NEXT/1 machine*) OR big-data OR linked-data OR (data NEAR/6 correlation*) OR (automat* NEAR/6 data NEAR/6 (analy* OR cluster* OR process*)) OR mapreduce OR (semantical* NEAR/6 text NEAR/6 cluster*)):ab,ti) AND ('editorial'/de OR (editorial*):ab,ti)

Medline Ovid

(exp "Machine Learning"/ OR "Pattern Recognition, Automated"/ OR "Automatic Data Processing"/ OR (("Data Interpretation, Statistical"/) AND ("automation"/ OR "Computational Biology"/ OR "Medical Informatics"/ OR "Biomedical Technology"/)) OR (((machine OR Bayesian OR network OR Autonom* OR semi-supervis* OR semisupervis* OR unsupervis*) ADJ3 learning) OR data-mining OR text-mining OR ((automat* OR algorithm* OR bioinformat*) ADJ6 (pattern* OR speech OR parameter*) ADJ6 (recogni* OR select*)) OR ((classif* OR learning) ADJ6 (algorithm* OR automat*)) OR (computer* ADJ3 heuristic*) OR (knowledge* ADJ3 discover*) OR (online ADJ6 analytic* ADJ6 process*) OR (vector ADJ machine*) OR big-data OR linked-data OR (data ADJ6 correlation*) OR (automat* ADJ6 data ADJ6 (analy* OR cluster* OR process*)) OR mapreduce OR (semantical* ADJ6 text ADJ6 cluster*)).ab,ti,kf.) AND ("editorial".pt. OR (editorial*).ab,ti,kf.)

Web of science

TS = (((((machine OR Bayesian OR network OR Autonom* OR semi-supervis* OR semisupervis* OR unsupervis*) NEAR/2 learning) OR data-mining OR text-mining OR ((automat* OR algorithm* OR bioinformat*) NEAR/5 (pattern* OR speech OR parameter*) NEAR/5 (recogni* OR select*)) OR ((classif* OR learning) NEAR/5 (algorithm* OR automat*)) OR (computer* NEAR/2 heuristic*) OR (knowledge* NEAR/2 discover*) OR (online NEAR/5 analytic* NEAR/5 process*) OR (vector NEAR/1 machine*) OR "Big Data" OR "linked data" OR (data NEAR/5 correlation*) OR (automat* NEAR/5 data NEAR/5 (analy* OR cluster* OR process*)) OR mapreduce OR (semantical* NEAR/5 text NEAR/5 cluster*))) AND (health* OR medicine* OR hospital* OR patient* OR disease* OR diagnos* OR therap* OR disorder*)) AND (DT = ("editorial material") OR TS = (editorial*))

Scopus

TITLE-ABS-KEY(((((machine OR Bayesian OR network OR Autonom* OR semi-supervis* OR semisupervis* OR unsupervis*) W/2 learning) OR data-mining OR text-mining OR ((automat* OR algorithm* OR bioinformat*) W/5 (pattern* OR speech OR parameter*) W/5 (recogni* OR select*)) OR ((classif* OR learning) W/5 (algorithm* OR automat*)) OR (computer* W/2 heuristic*) OR (knowledge* W/2 discover*) OR (online W/5 analytic* W/5 process*) OR (vector W/1 machine*) OR "Big Data" OR "linked data" OR (data W/5 correlation*) OR (automat* W/5 data W/5 (analy* OR cluster* OR process*)) OR mapreduce OR (semantical* W/5 text W/5 cluster*))) AND (health* OR medicine* OR hospital* OR patient* OR disease* OR diagnos* OR therap* OR disorder*)) AND (DOCTYPE(Editorial) OR TITLE-ABS-KEY(editorial*))

LISTA EBSCOhost

(MH "Machine Learning+" OR MH "Big Data+" OR MH "Data mining+" OR MH "electronic data processing" OR ((MH "Data analysis") AND (MH "automation" OR MH "Computer algorithms+" OR mh "Medical Informatics")) OR TI (((machine OR Bayesian OR network OR Autonom* OR semi-supervis* OR semisupervis* OR unsupervis*) N2 learning) OR data-mining OR text-mining OR ((automat* OR algorithm* OR bioinformat*) N5 (pattern* OR speech OR parameter*) N5 (recogni* OR select*)) OR ((classif* OR learning) N5 (algorithm* OR automat*)) OR (computer* N2 heuristic*) OR (knowledge* N2 discover*) OR (online N5 analytic* N5 process*) OR (vector N1 machine*) OR big-data OR linked-data OR (data N5 correlation*) OR (automat* N5 data N5 (analy* OR cluster* OR process*)) OR mapreduce OR (semantical* N5 text N5 cluster*)) OR AB (((machine OR Bayesian OR network OR Autonom* OR semi-supervis* OR semisupervis* OR unsupervis*) N2 learning) OR data-mining OR text-mining OR ((automat* OR algorithm* OR bioinformat*) N5 (pattern* OR speech OR parameter*) N5 (recogni* OR select*)) OR ((classif* OR learning) N5 (algorithm* OR automat*)) OR (computer* N2 heuristic*) OR (knowledge* N2 discover*) OR (online N5 analytic* N5 process*) OR (vector N1 machine*) OR big-data OR linked-data OR (data N5 correlation*) OR (automat* N5 data N5 (analy* OR cluster* OR process*)) OR mapreduce OR (semantical* N5 text N5 cluster*))) AND (PT "editorial" OR TI (editorial*) OR AB (editorial*)) AND (MH "medical records+" OR MH "medicine+" OR MH "medical informatics+" OR TI (health* OR medicine* OR hospital* OR patient* OR disease* OR diagnos* OR therap* OR disorder*) OR AB (health* OR medicine* OR hospital* OR patient* OR disease* OR diagnos* OR therap* OR disorder*))

Google Scholar

"machine|Bayesian|network|Autonomous learning"|"data mining"|"automated pattern|speech recognition"|"classification|learning algorithm"|"vector machine"|"big|linked data"|mapreduce intitle:editorial health|medicine|hospital|patient|diseases

Appendix 2: List of included editorials

Abbott CC, Loo D and Sui J (2016) Determining electroconvulsive therapy response with machine learning. *JAMA Psychiatry* 73(6): 545.

Ackland GL and Stephens RCM (2016) Big Data: A cheerleader for translational perioperative medicine. *Anesthesia & Analgesia* 122(6): 1744–1747.

Ahmad T, Testani JM and Desai NR (2016) Can Big Data simplify the complexity of modern medicine?: Prediction of right ventricular failure after left ventricular assist device support as a test case. *JACC. Heart failure* 4(9): 722–725.

Al Kazzi ES and Hutfless S (2015) Better Big Data. *Expert Review of Pharmacoeconomics & Outcomes Research* 15(6): 873–876.

Allarakhia M (2014) The successes and challenges of open-source biopharmaceutical innovation. *Expert Opinion on Drug Discovery* 9(5): 459–465.

Alter DA (2015) Merits and pitfalls of using observational 'Big Data' to inform our understanding of socioeconomic outcome disparities. *Journal of the American College of Cardiology* 66(17): 1898–1900.

Altman RB and Ashley EA (2015) Using 'Big Data' to dissect clinical heterogeneity. *Circulation* 131(3): 232–233.

Amato F, López A, Peña-Méndez EM, et al. (2013) Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine* 11(2): 47–58.

Arima H (2016) Utilizing Big Data for public health. *Journal of Epidemiology* 26(3): 105–105.

Atun R, Lussier Y, Poon C, et al. (2015) Editorial: Big Data for health. *IEEE Journal of Biomedical and Health Informatics* 19(4): 1191–1192.

Bagshaw SM, Goldstein SL, Ronco C, et al. (2016) Acute kidney injury in the era of Big Data: The 15th consensus conference of the Acute Dialysis Quality

Initiative (ADQI). *Canadian Journal of Kidney Health and Disease* 3(5): 103.

Baines D (2013) Big Data: Not just a lot more data. *Prescriber* 24(13–16): 7–8.

Barton AJ (2016) Big Data. *Journal of Nursing Education* 55(3): 123–124.

Basak SC, Vracko M and Bhattacharjee AK (2015) Big Data and new drug discovery: Tackling 'Big Data' for virtual screening of large compound databases. *Current Computer-Aided Drug Design* 11(3): 197–201.

Beck AH (2015) Open access to large scale datasets is needed to translate knowledge of cancer heterogeneity into better patient outcomes. *PLOS Medicine* 12(2): e1001794.

Belgrave D, Simpson A and Custovic A (2014) Challenges in interpreting wheeze phenotypes: The clinical implications of statistical learning techniques. *American Journal of Respiratory and Critical Care Medicine* 189(2): 121–123.

Benedict SH, El Naqa I and Klein EE (2016) Introduction to Big Data in radiation oncology: Exploring opportunities for research, quality assessment, and clinical care. *International Journal of Radiation Oncology*Biology*Physics* 95(3): 871–872.

Blankenship JC (2015) The power of the story in an era of Big Data and huge databases. *The Journal of Invasive Cardiology* 27(1): 33–34.

Boer C (2016) Scientific research in the perioperative period: Facing a changing (digital) context. *Netherlands Journal of Critical Care* 24(1): 4–5.

Boland MV (2016) Big Data, big challenges. *Ophthalmology* 123(1): 7–8.

Bourne PE (2014) What Big Data means to me. *Journal of the American Medical Informatics Association* 21(2): 194–194.

Bourne PE (2015) Confronting the ethical challenges of Big Data in public health. *PLOS Computational Biology* 11(2): e1004073.

Bourne PE, Bonazzi V, Dunn M, et al. (2015) The NIH Big Data to knowledge (BD2K) initiative. *Journal of the American Medical Informatics Association* 22(6): 1114–1114.

Brahmachari SK (2012) Introducing the medical bioinformatics in *Journal of Translational Medicine*. *Journal of Translational Medicine* 10(1): 202.

Brawley OW (2016) The cancer registry as a cancer–control tool: Cancer Registry. *Cancer* 122(9): 1343–1345.

Broome ME (2016) Big Data, data science, and big contributions. *Nursing Outlook* 64(2): 113–114.

Brown ML (2016) Can't you just pull the data? The limitations of using of the electronic medical record for research. *Pediatric Anesthesia* 26(11): 1034–1035.

Caban JJ and Gotz D (2015) Visual analytics in healthcare – Opportunities and research challenges. *Journal of the American Medical Informatics Association* 22(2): 260–262.

Cases M, Furlong LI, Albanell J, et al. (2013) Improving data and knowledge management to better integrate health care and research. *Journal of Internal Medicine* 274(4): 321–328.

Celi LA, Mark RG, Stone DJ, et al. (2013) 'Big Data' in the intensive care unit. Closing the data loop. *American Journal of Respiratory and Critical Care Medicine* 187(11): 1157–1160.

Chang AC and Hunt J (2014) Toward unreasonable effectiveness of cardiac ICU data: Artificial intelligence in pediatric cardiac intensive care*. *Pediatric Critical Care Medicine* 15(6): 565–567.

Chang H and Choi M (2016) Big Data and healthcare: Building an augmented world. *Healthcare Informatics Research* 22(3): 153.

Chen J (2016) Trying to understand nonarteritic anterior ischemic optic neuropathy through Big Data. *Ophthalmology* 123(12): 2442–2443.

Chen RC, Gabriel PE, Kavanagh BD, et al. (2016) How will Big Data impact clinical decision making and precision medicine in radiation therapy? *International Journal of Radiation Oncology*Biology*Physics* 95(3): 880–884.

Clancy TR and Reed L (2016) Big Data, big challenges: Implications for chief nurse executives. *JONA: The Journal of Nursing Administration* 46(3): 113–115.

Coghill D (2015) Editorial: Painting by numbers: Using modern approaches to analyse and visualise clinical and research data. *Journal of Child Psychology and Psychiatry* 56(10): 1035–1037.

Cooper D, Limet N, McClung I, et al. (2013) Towards clinically useful neuroimaging in psychiatric practice. *The British Journal of Psychiatry* 203(4): 242–244.

Craven M and Page CD (2015) Big Data in healthcare: Opportunities and challenges. *Big Data* 3(4): 209–210.

Crump C, Sundquist K and Winkleby MA (2015) Transnational research partnerships: Leveraging big data to enhance US health. *Journal of Epidemiology and Community Health* 69(11): 1029–1030.

DeKosky ST (2014) The role of Big Data in understanding late-life cognitive decline: E Unum, Pluribus. *JAMA Neurology* 71(12): 1476.

De Lemos JA, Rohatgi A and Ayers CR (2015) Applying a Big Data approach to biomarker discovery: Table: Running before we walk? *Circulation* 132(24): 2289–2292.

Denaxas SC, Asselbergs FW and Moore JH (2016) The tip of the iceberg: Challenges of accessing hospital electronic health record data for biological data mining. *BioData Mining* 9(1): 29.

Dereli T, Coşkun Y, Kolker E, et al. (2014) Big Data and ethics review for health systems research in LMICs: Understanding risk, uncertainty and ignorance – And catching the black swans? *The American Journal of Bioethics* 14(2): 48–50.

Dhar V (2014) Big Data and predictive analytics in health care. *Big Data* 2(3): 113–116.

Doll JA and Patel MR (2015) Self-regulation in the era of Big Data: Appropriate use of appropriate use criteria. *Annals of Internal Medicine* 162(8): 592.

Egan BM (2013) Prediction of incident hypertension. Health implications of data mining in the 'Big Data' era. *Journal of Hypertension* 31(11): 2123–2124.

Estes EH (2015) Big Data and the electrocardiogram. *Journal of Electrocardiology* 48(1): 29–30.

Eytan D, Goodwin A, Laussen P, et al. (2016) Insights from multi-dimensional physiological signals to predict and prevent cardiac arrests*. *Pediatric Critical Care Medicine* 17(1): 81–82.

Flockhart D, Bies RR, Gastonguay MR, et al. (2016) Big Data: Challenges and opportunities for clinical pharmacology. *British Journal of Clinical Pharmacology* 81(5): 804–806.

Fodeh S and Zeng Q (2016) Mining Big Data in biomedicine and health care. *Journal of Biomedical Informatics* 63: 400–403.

Freeman JV and Saxon L (2015) Remote monitoring and outcomes in pacemaker and defibrillator patients. *Journal of the American College of Cardiology* 65(24): 2611–2613.

Fridley BL, Koeslter DC and Godwin AK (2014) Individualizing care for ovarian cancer patients using Big Data. *JNCI: Journal of the National Cancer Institute* 106(5): 1–2.

Gaile DP and Miecznikowski JC (2013) From small studies to precision medicine: Prioritizing candidate biomarkers. *Genome Medicine* 5(11): 104.

Garrard P and Elvevåg B (2014) Language, computers and cognitive neuroscience. *Cortex* 55: 1–4.

Garrison LP (2013) Universal health coverage – Big thinking versus Big Data. *Value in Health* 16(1): S1–S3.

Ghani KR, Zheng K, Wei JT, et al. (2014) Harnessing Big Data for health care and research: Are urologists ready? *European Urology* 66(6): 975–977.

Giambrone GP, Hemmings HC, Sturm M, et al. (2015) Information technology innovation: The power and perils of big data. *British Journal of Anaesthesia* 115(3): 339–342.

Gomella LG (2016) Big Data equals big challenges for prostate cancer. *The Canadian Journal of Urology* 23(4): 8329.

Goozner M (2013) Big Data and big government. Strong federal role needed to organize productive use of patient data. *Modern Healthcare* 43(25): 16.

Goozner M (2014) Better rules needed to boost use of Big Data. *Modern Healthcare* 44(28): 38.

Grauer JN and Leopold SS (2015) Editorial: Large database studies – What they can do, what they cannot do, and which ones we will publish. *Clinical Orthopaedics and Related Research* 473(5): 1537–1539.

Green DE and Rapp EJ (2013) Can Big Data lead us to big savings? *RadioGraphics* 33(3): 859–860.

Groeneveld PW and Rumsfeld JS (2016) Can Big Data fulfill its promise? *Circulation: Cardiovascular Quality and Outcomes* 9(6): 679–682.

Hauptman PJ (2016) Got Big Data? *Journal of Cardiac Failure* 22(3): 169–170.

Hay SI, George DB, Moyes CL, et al. (2013) Big Data opportunities for global infectious disease surveillance. *PLoS Medicine* 10(4): 1–4.

Helfet DL, Hanson BP and De Faoite D (2014) Big Data: The paradigm shift needed to revolutionize musculoskeletal clinical research. *American Journal of Orthopedics* 43(9): 399–400.

Henly SJ (2014) Mother lodes and mining tools: Big Data for nursing science. *Nursing Research* 63(3): 155.

Hochster HS and Niedzwiecki D (2016) Big Data, small effects. *Journal of Clinical Oncology* 34(11): 1170–1171.

Hoffman S (2016) The promise and perils of open medical data. *Hastings Center Report* 46(1): 6–7.

Holland CM, Foley KT and Asher AL (2015) Editorial. Can Big Data bridge the chasm? Issues, opportunities, and strategies for the evolving value-based health care environment. *Neurosurgical Focus* 39(6): E2.

Horvitz E and Mulligan D (2015) Data, privacy, and the greater good. *Science* 349(6245): 253–255.

Huilgol N (2016) Big Data in radiation oncology. *Journal of Cancer Research and Therapeutics* 12(3): 1107–1108.

Hunt J and Chang A (2013) Big Data in pediatric cardiac care: Is it time?*. *Pediatric Critical Care Medicine* 14(5): 548–549.

Husain M (2014) Big Data: Could it ever cure Alzheimer's disease? *Brain* 137(10): 2623–2624.

Huser V and Cimino JJ (2016) Impending challenges for the use of Big Data. *International Journal of Radiation Oncology*Biology*Physics* 95(3): 890–894.

Hu Y and Bajorath J (2014) Learning from 'Big Data': Compounds and targets. *Drug Discovery Today* 19(4): 357–360.

Issa AM, Marchant GE and Campos-Outcalt D (2016) Big Data in the era of precision medicine: big promise or big liability? *Personalized Medicine* 13(4): 283–285.

Jiang H, An L, Baladandayuthapani V, et al. (2014) Classification, predictive modelling, and statistical analysis of cancer data (A). *Cancer Informatics* 13s2: 1–3.

Joynt KE, Mega JL and O'Donoghue ML (2015) Difference or disparity: Will Big Data improve our understanding of sex and cardiovascular disease? *Circulation: Cardiovascular Quality and Outcomes* 8(2): S52–S55.

Kankanhalli A, Hahn J, Tan S, et al. (2016) Big Data and analytics in healthcare: Introduction to the special section. *Information Systems Frontiers* 18(2): 233–235.

Kaufmann SHE, Fletcher HA, Guzmán CA, et al. (2015) Big Data in vaccinology: Introduction and section summaries. *Vaccine* 33(40): 5237–5240.

Khoury MJ (2015) Planning for the future of epidemiology in the era of Big Data and precision medicine: Table 1. *American Journal of Epidemiology* 182(12): 977–979.

Khoury MJ and Ioannidis JPA (2014) Big Data meets public health. *Science* 346(6213): 1054–1055.

Kieling C (2015) Here/In this issue and there/abstract thinking: Randomized controlled trials in the era of Big Data. *Journal of the American Academy of Child & Adolescent Psychiatry* 54(12): 967–968.

King SB (2016) Big trials or Big Data. *JACC: Cardiovascular Interventions* 9(8): 869–870.

Klonoff D (2013) Twelve modern digital technologies that are transforming decision making for diabetes and all areas of health care. *Journal of Diabetes Science and Technology* 7(2): 291–296.

Klonoff DC (2015) Precision medicine for managing diabetes. *Journal of Diabetes Science and Technology* 9(1): 3–7.

Kottyan LC, Weirauch MT and Rothenberg ME (2015) Making it big in allergy. *Journal of Allergy and Clinical Immunology* 135(1): 43–45.

Kovacs RJ (2015) Using practice-based evidence to reduce disparities in care. *Journal of the American College of Cardiology* 66(11): 1234–1235.

Krakoff LR and Phillips RA (2016) Blood pressure variability. *Journal of the American College of Cardiology* 68(13): 1387–1388.

Krumholz HM (2016) The promise of Big Data: Opportunities and challenges. *Circulation: Cardiovascular Quality and Outcomes* 9(6): 616–617.

Kuchenreuther M and Sackman JE (2014) Marrying Big Data with personalized medicine. *The Science & Business of Biopharmaceuticals* 27(8): 36–38.

Lee CS, Lee AY, Holland GN, et al. (2016) Big Data and uveitis. *Ophthalmology* 123(11): 2273–2275.

Leem JG (2016) Big Data and pain. *The Korean Journal of Pain* 29(4): 215.

Lentine KL and Segev DL (2013) Better understanding live donor risk through Big Data. *Clinical Journal of the American Society of Nephrology* 8(10): 1645–1647.

Liebeskind DS (2015) Big and bigger data in endovascular stroke therapy. *Expert Review of Neurotherapeutics* 15(4): 335–337.

Litman RS (2013) Complications of laryngeal masks in children: Big Data comes to pediatric anesthesia. *Anesthesiology* 119(6): 1239–1240.

Litman RS (2016) The use of patient registries to detect risk factors of anesthesia and sedation complications. *Pediatrics* 137(3): e20154579.

Liu B (2014) Utilizing Big Data to build personalized technology and system of diagnosis and treatment in traditional Chinese medicine. *Frontiers of Medicine* 8(3): 272–278.

Loder E and Burch R (2014) What can data mining teach us about triptan safety that we don't already know? *Cephalalgia* 34(1): 3–4.

Lusher SJ and Ritschel T (2015) Finding the right approach to Big Data-driven medicinal chemistry. *Future Medicinal Chemistry* 7(10): 1213–1216.

Lushington G (2014) Editorial: Chemical screening: Thinking big with Big Data. *Combinatorial Chemistry & High Throughput Screening* 17(6): 483–484.

Lynch SM and Moore JH (2016) A call for biological data mining approaches in epidemiology. *BioData Mining* 9(1): 1–3.

Mackie P, Sim F and Johnman C (2015) Big data! Big deal? *Public Health* 129(3): 189–190.

MacLaren G, Fortenberry JD and Dalton HJ (2016) Lies, statistics, and ECMO data mining: Digging dirt or striking gold?*. *Pediatric Critical Care Medicine* 17(8): 799–802.

MacRae CA (2012) Pattern recognition: Combining informatics and genetics to re-evaluate conduction disease. *Heart* 98(17): 1263–1264.

Madabhushi A and Lee G (2016) Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical Image Analysis* 33: 170–175.

Maddox TM and Matheny MA (2015) Natural language processing and the promise of Big Data: Small step forward, but many miles to go. *Circulation: Cardiovascular Quality and Outcomes* 8(5): 463–465.

Malley JD and Moore JH (2014) First complex, then simple. *BioData Mining* 7(1): 13.

Mardis ER (2016) The challenges of Big Data. *Disease Models & Mechanisms* 9(5): 483–485.

Martinez-Garcia MA and Dinh-Xuan AT (2016) Deriving information from external big databases and Big Data analytics: All that glitters is not gold. *European Respiratory Journal* 47(4): 1047–1049.

McDermott S and Turk MA (2015) What are the implications of the big data paradigm shift for disability and health? *Disability and Health Journal* 8(3): 303–304.

McNutt TR, Moore KL and Quon H (2016) Needs and challenges for Big Data in radiation oncology.

*International Journal of Radiation Oncology*Biology*Physics* 95(3): 909–915.

Mensah E (2016) Editorial Vol 8 Issue 3. *Online Journal of Public Health Informatics* 8(3): e194.

Merelli I, Pérez-Sánchez H, Gesing S, et al. (2014) High-performance computing and Big Data in omics-based medicine. *BioMed Research International* 2014: 1–2.

Morley S (2015) Big Data: Little difference. *European Journal of Pain* 19(5): 593–594.

Morris PJ (2014) The dawn of Big Data. *North Carolina Medical Journal* 75(3): 177.

Müller H, Hanbury A and Al Shorbaji N (2012) Health information search to deal with the exploding amount of health information produced. *Methods of Information in Medicine* 51(6): 516–518.

Murthy SC and Blackstone EH (2016) Research based on Big Data: The good, the bad, and the ugly. *The Journal of Thoracic and Cardiovascular Surgery* 151(3): 629–630.

Nadkarni GN and Coca SG (2016) Temporal trends in AKI: Insights from Big Data. *Clinical Journal of the American Society of Nephrology* 11(1): 1–3.

Nagueh SF (2016) Unleashing the potential of machine-based learning for the diagnosis of cardiac diseases. *Circulation: Cardiovascular Imaging* 9(6): e005059.

Najarian K, Ward KR and Shirani S (2013) Biomedical signal and image processing for clinical decision support systems. *Computational and Mathematical Methods in Medicine* 2013: 1–2.

Nallamothu BK (2016) Moving from Big Data to vital insights. *Circulation: Cardiovascular Quality and Outcomes* 9(6): 615–615.

Nandi U, Puskarich MA and Jones AE (2016) Big Data to the rescue of systemic inflammatory response syndrome: Is electronic data mining the way of the future? *Annals of Translational Medicine* 4(23): 465–465.

Narula J (2013) Are we up to speed? *JACC: Cardiovascular Imaging* 6(11): 1222–1224.

Nash D (2014) Harnessing the power of Big Data in healthcare. *American Health & Drug Benefits* 7(2): 69–70.

Nature (2016) The power of Big Data must be harnessed for medical progress. *Nature* 539(7630): 467–468.

Nature Chemical Biology (2014) Microbiota meet Big Data. *Nature Chemical Biology* 10(8): 605–605.

Nature Genetics (2013) Ministry of noise. *Nature Genetics* 45(8): 843.

Ndiaye NC (2014) Systems medicine in the era of ‘Big Data’: A game-changer for personalized medicine? *Drug Metabolism and Drug Interactions* 29(3).

Nicholls SJ and Psaltis PJ (2016) Will Big Data shine light at the end of the tunnel for HDL?∗. *Journal of the American College of Cardiology* 68(19): 2084–2085.

Ohno-Machado L and Editor-in-Chief (2015) Mining electronic health record data: Finding the gold nuggets. *Journal of the American Medical Informatics Association* 22(5): 937–937.

Oktar N and Oktar Y (2015) Machine learning and neuroimaging. *Journal of Neurological Sciences* 32(1): 1–4.

Onukwugha E (2016) Big Data and its role in health economics and outcomes research: A collection of perspectives on data sources, measurement, and analysis. *PharmacoEconomics* 34(2): 91–93.

Padrez KA, Asch DA and Merchant RM (2015) The patient diarist in the digital age. *Journal of General Internal Medicine* 30(6): 708–709.

Palumbo B and Fravolini ML (2012) To what extent can artificial neural network support nuclear medicine? *Hellenic Journal of Nuclear Medicine* 15(3): 180–183.

Parker RI (2015) Hematopoietic stem cell transplantation: Better data, better care? Big data, bigger questions!*. *Critical Care Medicine* 43(9): 2037–2038.

Patel PM and Chen Z (2012) When the doctor needs an engineer to be the matchmaker. *EuroIntervention: Journal of EuroPCR in Collaboration with the Working Group on Interventional Cardiology of the European Society of Cardiology* 8(1): 19–23.

Pathak J, Kho AN and Denny JC (2013) Electronic health records-driven phenotyping: Challenges, recent advances, and perspectives. *Journal of the American Medical Informatics Association* 20(e2): e206–e211.

Patrick K (2016) Harnessing Big Data for health. *Canadian Medical Association Journal* 188(8): 555–555.

Patt D (2016) Better screening using Big Data. *Journal of Oncology Practice* 12(8): 699–700.

Peek N and Abu-Hanna A (2014) Clinical prognostic methods: Trends and developments. *Journal of Biomedical Informatics* 48: 1–4.

Perry DC, Parsons N and Costa ML (2014) ‘Big Data’ reporting guidelines: How to answer big questions, yet avoid big problems. *Bone & Joint Journal* 96(12): 1575–1577.

Pezzi CM (2014) Big Data and clinical research in oncology: The good, the bad, the challenges, and the opportunities. *Annals of Surgical Oncology* 21(5): 1506–1507.

Pope C, Halford S, Tinati R, et al. (2014) What’s the big fuss about ‘big data’? *Journal of Health Services Research & Policy* 19(2): 67–68.

Porche DJ (2014) Men’s health Big Data. *American Journal of Men’s Health* 8(3): 189–189.

Potters L, Ford E, Evans S, et al. (2016) A systems approach using Big Data to improve safety and quality in radiation oncology. *International Journal of Radiation Oncology*Biology*Physics* 95(3): 885–889.

Ranney ML and Genes N (2016) Social media and healthcare quality improvement: A nascent field. *BMJ Quality & Safety* 25(6): 389–391.

Redline S, Dean D and Sanders MH (2013) Entering the era of 'Big Data': Getting our metrics right. *Sleep* 36(4): 465–469.

Restifo NP (2013) A 'Big Data' view of the tumor 'immunome'. *Immunity* 39(4): 631–632.

Riley D and Mittelman M (2012) Maps, 'Big Data,' and case reports. *Global Advances in Health and Medicine* 1(3): 5–7.

Risoud M, Bonne N-X and Vincent C (2016) Big Data: Coming soon to ENT. *European Annals of Otorhinolaryngology, Head and Neck Diseases* 133(3): 157.

Rodeghero J and Cook C (2014) The use of Big Data in manual physiotherapy. *Manual Therapy* 19(6): 509–510.

Rosenbloom ST (2016) Person-generated health and wellness data for health care. *Journal of the American Medical Informatics Association* 23(3): 438–439.

Rosenstein BS, West CM, Bentzen SM, et al. (2014) Radiogenomics: Radiobiology enters the era of Big Data and team science. *International Journal of Radiation Oncology*Biology*Physics* 89(4): 709–713.

Rueckert D, Glocker B and Kainz B (2016) Learning clinically useful information from images: Past, present and future. *Medical Image Analysis* 33: 13–18.

Rusconi B and Warner BB (2017) The hidden treasure of neonatal screening: Identifying new risk factors and possible mechanisms of necrotizing enterocolitis through Big Data. *The Journal of Pediatrics* 181: 9–11.

Sacristán JA and Dilla T (2015) No Big Data without small data: Learning health care systems begin and end with the individual patient: No Big Data without small data. *Journal of Evaluation in Clinical Practice* 21(6): 1014–1017.

Salcido R (2013) Big Data and disruptive innovation in wound care. *Advances in Skin & Wound Care* 26(8): 344.

Şardaş S, Endrenyi L, Gürsoy UK, et al. (2014) A call for pharmacogenovigilance and rapid falsification in the age of Big Data: Why not first road test your biomarker? *OMICS: A Journal of Integrative Biology* 18(11): 663–665.

Sarin R (2014) Big Data V4 for integrating patient reported outcomes and quality-of-life indices in clinical practice. *Department of Radiation Oncology* 10(3): 453–455.

Schadt EE (2012) The changing privacy landscape in the era of Big Data. *Molecular Systems Biology* 8: 612.

Schoenfeld AJ (2016) Research using 'Big Data' in orthopaedic trauma: A dynasty of databases or finite research resource? *Journal of Orthopaedic Trauma* 30(5): 225–227.

Schoenhagen P and Mehta N (2016) Big Data, smart computer systems, and doctor–patient relationship. *European Heart Journal* 38(7): 508–510.

Schubert M and Iorio F (2014) Exploiting combinatorial patterns in cancer genomic data for personalized therapy and new target discovery. *Pharmacogenomics* 15(16): 1943–1946.

Sessler DI (2014) Big Data – And its contributions to peri-operative medicine. *Anaesthesia* 69(2): 100–105.

Sethi A (2012) The future of healthcare informatics: It is not what you think. *Global Advances in Health and Medicine* 1(4): 5–6.

Shah NH and Tenenbaum JD (2012) The coming age of data-driven medicine: Translational bioinformatics' next frontier. *Journal of the American Medical Informatics Association* 19(1): e2–e4.

Sim I (2016) Two ways of knowing: Big Data and evidence-based medicine. *Annals of Internal Medicine* 164(8): 562–563.

Sørensen H, Langan, Benchimol, et al. (2013) Setting the RECORD straight: Developing a guideline for the reporting of studies conducted using observational routinely collected data. *Clinical Epidemiology* 5: 29–31.

Stohler C (2014) The next frontier: Digital disease detection in cyberspace. *Journal of Oral & Facial Pain and Headache* 28(2): 105.

Suominen H (2014) Text mining and information analysis of health documents. *Artificial Intelligence in Medicine* 61(3): 127–130.

Tajik AJ (2016) Machine learning for echocardiographic imaging. *Journal of the American College of Cardiology* 68(21): 2296–2298.

Tan SS, Gao G and Koch S (2015) Big Data and analytics in healthcare. *Methods of Information in Medicine* 54(6): 546–547.

Thorpe JH and Gray EA (2015) Big Data and public health: Navigating privacy laws to maximize potential. *Public Health Reports* 130(2): 171–175.

Tilson H (2015) The evidence for a changing real world of real world evidence. *Current Medical Research and Opinion* 31(5): 1027–1028.

Tong H, Wang F, Choudhury MD, et al. (2016) Guest Editorial: Special issue on connected health at Big Data era (BigChat): A TKDD special issue. *ACM Transactions on Knowledge Discovery from Data* 10(4): 1–4.

Trägårdh E, Carlsson M and Edenbrandt L (2015) Computerized decision making in myocardial perfusion SPECT: The new era in nuclear cardiology? *Journal of Nuclear Cardiology* 22(5): 885–887.

Tsalik EL (2014) Crowdsourcing disease prognosis. *Science Translational Medicine* 6(264): 264ec204–264ec204.

Valenstein M (2013) The promise of large, longitudinal data sets. *Psychiatric Services* 64(6): 503–503.

Valentini V and Cellini C (2014) New perspectives in treatment decision for integrated management of rectal cancer: Multimodal research for multimodal treatments. *Journal of Surgery* 35(5–6): 113–116.

Van De Ville D and Lee SW (2012) Brain decoding: Opportunities and challenges for pattern recognition. *Pattern Recognition* 45(6): 2033–2034.

Van Hooijdonk RTM, Krinsley JS and Schultz MJ (2016) DETECT the extremes that usually remain undetected in conventional observational studies. *Clinical Chemistry* 62(5): 668–670.

Vayena E, Salathé M, Madoff LC, et al. (2015) Ethical challenges of Big Data in public health. *PLOS Computational Biology* 11(2): e1003904.

Vetter TR and Redden DT (2015) The power and perils of Big Data: It all depends on how you slice, dice, and digest it. *Anesthesia & Analgesia* 121(3): 582–585.

Von Gunten CF, Teno JM and Morrison R. (2016) Big Data and end-of-life care: Promise and peril. *Journal of Palliative Medicine* 19(12): 1240–1240.

Waldman S and Terzic A (2016) Big Data transforms discovery-utilization therapeutics continuum. *Clinical Pharmacology & Therapeutics* 99(3): 250–254.

Waller LA and Miller GW (2016) More than manuscripts: Reproducibility, rigor, and research productivity in the Big Data era. *Toxicological Sciences* 149(2): 275–276.

Wang L and Xie X-Q (2016) Cancer genomics: opportunities for medicinal chemistry? *Future Medicinal Chemistry* 8(4): 357–359.

Wang L, Ranjan R, Kołodziej J, et al. (2015) Software tools and techniques for Big Data computing in healthcare clouds. *Future Generation Computer Systems* 43–44: 38–39.

Wang SD (2013) Opportunities and challenges of clinical research in the Big-Data era: From RCT to BCT. *Journal of Thoracic Disease* 5(6): 721–723.

Wang SD and Shen Y (2015) Big-Data clinical trial (BCT): The third talk. *Journal of Thoracic Disease* 7(8): E243–E244.

Wasfy JH and Maddox TM (2016) Worlds apart: The hype and reality of Big Data to improve health care. *Circulation: Cardiovascular Quality and Outcomes* 9(5): 495–497.

Webster M and Kumar VS (2014) Big Data diagnostics. *Clinical Chemistry* 60(8): 1130–1132.

Weil AR (2014) Big Data in health: A new era for research and patient care. *Health Affairs* 33(7): 1110–1110.

Weinstein JN (2016) An 'Industrial Revolution' in health care: The data tell us the time has come. *Spine* 41(1): 1–2.

Wiederhold BK (2012) Self-tracking: Better medicine through pattern recognition. *Cyberpsychology, Behavior, and Social Networking* 15(5): 235–236.

Williams SM and Moore JH (2015) Lumping versus splitting: The need for biological data mining in precision medicine. *BioData Mining* 8(1): 16.

Yang CC and Veltri P (2015) Intelligent healthcare informatics in Big Data era. *Artificial Intelligence in Medicine* 65(2): 75–77.

Yli-Hietanen J, Ylipää A and Yli-Harja O (2015) Cancer research in the era of next-generation sequencing and Big Data calls for intelligent modeling. *Chinese Journal of Cancer* 34(3): 12.

Young SD (2014) Behavioral insights on Big Data: Using social media for predicting biomedical outcomes. *Trends in Microbiology* 22(11): 601–602.

Zeng QT and Fodeh S (2015) Clinical data mining. *Computers in Biology and Medicine* 62: 293.

Zhang S and Metaxas D (2016) Large-scale medical image analytics: Recent methodologies, applications and future directions. *Medical Image Analysis* 33: 98–101.

Zhang Z (2016) A gentle introduction to artificial neural networks. *Annals of Translational Medicine* 4(19): 370–370.

Zhongheng Z (2014) Big Data and clinical research: Perspective from a clinician. *Journal of Thoracic Disease* 6(12): 1659–1664.

Zurlinden J (2016) Health professions education: Make Big Data meaningful. *The Journal of Continuing Education in Nursing* 47(5): 199–200.