

Big Data Analytics, Infectious Diseases and Associated Ethical Impacts

Chiara Garattini¹ · Jade Raffle² · Dewi N Aisyah³ ·
Felicity Sartain⁴ · Zisis Kozlakidis² 

Received: 24 February 2017 / Accepted: 2 August 2017 / Published online: 24 August 2017
© The Author(s) 2017. This article is an open access publication

Abstract The exponential accumulation, processing and accrual of big data in healthcare are only possible through an equally rapidly evolving field of big data analytics. The latter offers the capacity to rationalize, understand and use big data to serve many different purposes, from improved services modelling to prediction of treatment outcomes, to greater patient and disease stratification. In the area of infectious diseases, the application of big data analytics has introduced a number of changes in the information accumulation models. These are discussed by comparing the traditional and new models of data accumulation. Big data analytics is fast becoming a crucial component for the modelling of transmission—aiding infection control measures and policies—emergency response analyses required during local or international outbreaks. However, the application of big data analytics in infectious diseases is coupled with a number of ethical impacts. Four key areas are discussed in this paper: (i) automation and algorithmic reliance impacting freedom of choice, (ii) big data analytics complexity impacting informed consent, (iii) reliance on profiling impacting individual and group identities and justice/fair access and (iv) increased surveillance and population intervention capabilities impacting behavioural norms and practices. Furthermore, the extension of big data analytics to include information derived from personal devices, such as mobile phones and wearables as part of infectious disease frameworks in the near future and their potential ethical impacts are discussed. Considered together,

Chiara Garattini and Jade Raffle contributed equally to this work.

✉ Zisis Kozlakidis
z.kozlakidis@ucl.ac.uk

¹ Anthropology and UX Research, Health and Life Sciences, Intel, London, UK

² Division of Infection and Immunity, University College London, Cruciform Building, Gower Street, London WC1E 6BT, UK

³ Department of Infectious Disease Informatics, University College London, Farr Institute of Health Informatics Research, 222 Euston Road, London NW1 2DA, UK

⁴ Profiscio Ltd., 10 Great Russell Street, London WC1B 3BQ, UK

the need for a constructive and transparent inclusion of ethical questioning in this rapidly evolving field becomes an increasing necessity in order to provide a moral foundation for the societal acceptance and responsible development of the technological advancement.

Keywords Infectious diseases · Big data analytics · Ethics · Mobile phones · Wearables

1 Introduction

The terms big data and big data analytics are often used within the context of healthcare as an all-encompassing phrase referring to the use of large datasets. Their increasingly regular use does little to signify the underlying complexity of definitions, and paves the way for potential subsequent ethical and social misunderstandings (Florida 2012). Big data is defined as collections of data so large and highly complex that its manipulation and management require the application of a series of computing techniques—including but not limited to—machine learning and artificial intelligence (Stuart Ward and Barker 2013). Big data analytics is defined as ‘the process of collecting, organizing and analysing large sets of data (called big data) to discover patterns and other useful information’ (Heymann and Rodier 2004). Within the sphere of routine healthcare and associated research, the term is often synonymous to electronic patient records from central public authorities, large hospitals, clinical trials, ‘-omics’ and sequencing-based outputs and their associated banked samples, imaging, mobile phones and/or wearables. These data can be structured or unstructured, generated from diverse sources and at diverse speed, and in very large volumes. Big data analytics (BDA) are the collectively termed processes that translate the big data into interpretable and potentially actionable sets of information that can confer a competitive advantage (LaValle and Lesser 2011).

BDA have attracted attention as they have been shown to have transformational potential for generating reliable, actionable and novel insights about the world we live in. Positive examples have been reported in the financial services (Mayer-Schoenberger and Cukier 2013) and in healthcare (Murdoch and Detsky 2013; Raghupathi and Raghupathi 2014; Lee et al. 2016). For infectious diseases, BDA is useful in monitoring diseases outbreaks; stratifying patients for treatment, risk exposure and treatment outcome prediction; and for the prediction of behavioural patterns with the aim of informing public health interventions (Darrell et al. 2015; Lee et al. 2016). Despite these positive outcomes, the routine use of BDA is increasingly challenged with examples of wrongful conclusions (Lazer et al. 2014), potential misuse of personal information, well-publicised privacy breeches and ongoing profiling of individuals for commercial purposes (Smith 2012; Richards and King 2014; Martin 2015). Often described as unintended (Wigan and Clarke 2013), these are very real risks (Clarke 2016). However, these risks are not always predictable or preventable, as the context in which big data is used heavily determines which of those problems actually occur and whether they may lead to negative consequences at the individual and/or societal levels (Markus 2015; Zuboff 2015; Duhigg 2012).

In the field of medical ethics, the basic principles of autonomy, beneficence, non-maleficence and justice remain relevant for BDA (Gillon 1994; Page 2012).

However, they are tested by the differential interpretation relating to their contextual application in clinical care, public health policy, administrative repurposing, patient-generated data and research (e.g. clinical trials and/or use of metadata) with BDA acting as a magnifying glass for potential differences and difficulties (Bollier 2010). Furthermore, the way in which data have increasingly been collected outside of traditional healthcare settings and shared with third parties for research and commercial gains has also changed (Jetten and Sharon 2016). These factors are expected to intensify further as data collection is expected to scale up and become more distributed through ‘non-traditional’ channels with its processing being quicker and more automated.

Given these potential issues associated with BDA and considering its increased uptake, discussions are ongoing regarding what constitutes its ethical versus unethical use in healthcare. There are a number of studies addressing this aspect (Mittelstadt and Floridi 2016), which is particularly pertinent in the field of infectious diseases, where the impact of BDA is already tangible and is further augmented by the ‘non-traditional’ data collection systems, such as wearables and mobile phones (Vanheems et al. 2013). In this paper, we discuss some of the changes occurring in the field of BDA in infectious diseases by comparing the traditional model of data accumulation to the new one. We then discuss the ethical challenges according to four headings: (i) automation and algorithmic reliance impacting freedom of choice, (ii) big data analytics complexity impacting informed consent, (iii) reliance on profiling impacting individual and group identities and justice/fair access and (iv) increased surveillance and population intervention capabilities impacting behavioural norms and practices.

2 Information Accumulation Models in Infectious Diseases and the Impact of BDA

The traditional information accumulation model in infectious diseases remains a variant of a hierarchical hub-and-spoke model, where well-defined smaller reporting centres (such as GP practices or individual virologists) report to a central authority at the local level (such as referral hospitals) or at the national level (such as Public Health England in the UK) for the epidemiological view. The accumulated information is aggregated at population level, processed, corrected/defined and processed again (in a number of iterative cycles). Subsequently, actions relating to these conclusions are disseminated to the whole system in a top-down approach. This is economical in the sense of using defined communication channels according to established professional and well-regulated processes, recommendations, actors and actions. In addition, there are bona fide population data aggregators, such as national health surveys, administrative hospital data and biobanks, at a local national or international level, for example the UK Biobank in the UK and the BBMRI-ERIC in the European Union (Biobanking and Biomolecular Resources research infrastructure–European Research Infrastructure Consortium) (Kozlakidis 2016).

However, the information accumulation and processing steps carry an inevitable time lag that can result in reduced response effectiveness at a public health intervention level. The usually slow methods of assessing medical knowledge through the

accumulation of evidence-based cases are challenged by rapidly evolving global threats. This can have catastrophic results when dealing with infectious disease occurrences as was observed in the case of the recent Ebola outbreak in West Africa (McCarthy 2016) or in the cases of the global influenza pandemics in 2008–2009 (Doshi 2009). It should be noted though that the established, global hub-and-spoke system is still considered successful in identifying potential disease outbreaks at an international level (Heyman and Rodier 2001; Heyman and Rodier 2004) and some national levels (Randrianasolo 2010). The need for speed in data sharing/accumulation, processing and advice provision in the case of unexpected/novel outbreaks was identified early on and became one of the main drivers for introducing BDA capabilities. The initial piloted learnings were then distilled and used as improvements on the hub-and-spoke systems (Jacobsen et al. 2016; Simonsen et al. 2016).

The above information accumulation models are most effective in the cases of sedentary, well-defined, well-predictable populations and outbreaks. They do not necessarily reflect the current global contextual complexity or the patchy and/or often unstructured nature of real-time data. As such, they come under increasing stress and criticism for lack of adaptability. Mass migrations—often resulting from emergencies such as floods and wars—can produce conditions that exacerbate this phenomenon, spreading infectious diseases, often in unexpected ways. Global trade, mass gatherings and travel can introduce pathogens into new populations creating the potential to develop into epidemics, e.g. West Nile virus introduction and spread in New York (Fonkwo 2008), post-festival measles outbreak in Germany (Pfaff et al. 2010) and H1N1 spread through airline transportation (Khan et al. 2009).

2.1 New Sources of Information Accumulation

Mobile phones can be used to address global and healthcare data inequities and hold particular promise in low- and middle-income countries (LMIC) where conventional sources of social and health-related data are often patchy, out of date or simply non-existent (Kaplan 2006; Deville et al. 2014; Center for Global Development 2014). The extraordinary reach and potential effectiveness of engaging mobile technologies was demonstrated by (a) an emergency reporting system for infectious diseases surveillance in Sichuan province, China, following the catastrophic earthquake of 2008 (Yang et al. 2009); (b) a monitoring system for the long-distance observation of tuberculosis patients in Kenya (Hoffman et al. 2010) and (c) a guidance tool for potential Ebola patients providing directions to the closest available health centres (Trad et al. 2015). In the UK, a mobile phone-enabled, video-observed therapy clinical trial for adherence to treatment for tuberculosis patients is already underway (Story et al. 2016) and under consideration as part of an increased patient offering by the UK public health system. In addition, a number of mobile-enabled infectious diseases diagnostic tools are currently under development. The use of mobile phone technologies as part of mainstream healthcare is considered largely inevitable (Dell et al. 2011).

There are many studies describing the development of different sensor technologies and wireless medical instruments (collectively termed as wearables) with the ability to remotely monitor patients at a distance, at home or elsewhere, by recording personal parameters such as blood pressure, heart rate, sugar levels, vital signs, oxygen levels, body temperature, and mental well-being metrics

(Chana 2012). The main driver behind these developments remains the need for the early warnings for the onset of adverse health conditions and service improvements coupled with costs reductions in the health sector (Page 2015a). Wearable devices can transmit vast amounts of data sometimes in constant live streaming modes which raises important ethical issues regarding privacy and security. Individuals, organisations and businesses can be severely affected if third parties access private information in an unlawful manner. Although wearable devices are designed to promote independence, they require different levels of privacy intrusion to collect data, especially when recording passively, therefore making it important to consider ethical implications from different stakeholders' perspectives (Yeslam 2015). Additionally, choice/autonomy and security/privacy risks can be bidirectional, i.e. from pre-selected incoming data (i) equipment might be adjusted to allow for greater provider preference as opposed to patient preference, in the case of medication pre-order for example and (ii) risks can arise from the outgoing information, as well as by incoming information into the device, as malware can jeopardize patients' health by causing some (intentional or unintentional) malfunction of the device (Segura et al. 2017).

2.2 BDA Impact on Infectious Diseases

The most commonly mentioned impact of BDA in the field of infectious diseases relates to the access and use of information contained in individual medical records and/or collected using healthcare systems' channels. In this context, insights generated by routine access and use of these data, including when explicit consent each time is not a viable option (such as in the case of large-scale, multiple-site epidemiological studies), can potentially benefit patients, institutions, public health as well as commercial entities. The main difference with the traditional regime relates to the scale and reach of BDA capabilities, accessing and processing much larger numbers of medical records and more information per record. However, the levels of access and use of the information do vary between locations and for different purposes. In the field of infectious diseases, decision is guided by two separate specialties' viewpoints: epidemiology, looking at the overall population-level health, and infectious diseases specialties (such as microbiology, virology) addressing the individual-level health. This is reflected in the ethical arguments framed as autonomy versus public good, where BDA has acted as a catalyst in the arguments on either side (Gilbert 2012; Kozlakidis et al. 2012; Blais and White 2015).

From an individual perspective, the availability of vast amounts of information regarding people, for example via geo-tagged social networks, makes even data anonymization ineffective in fully protecting the identity of the data source, making it only more difficult, yet still feasible via triangulation, to (re)identify it (Cecaj et al. 2016). As such, the ethical imperative of transparency with regard to the dangers of downstream data linkage and inadvertent individual identification should be upheld. To this end, the Groupe Speciale Mobile Association (GSMA) has developed guidelines for the appropriate use of Call Detail Record (CDR) and Standard Messaging Services (SMS) data in emergency situations (GSMA 2013) and updated them specifically for infectious diseases based on the experiences

from the Ebola outbreak in West Africa (GSMA 2014) with some limited but not extensive consideration of potential ethical implications.

From a population-level perspective, if systems are designed to rely entirely on anonymous contributions in order to protect their original data contributors, they might not work well either, as the element of information accountability and, hence, transparency is affected. Especially in the case of humanitarian emergencies, and certainly communicable disease outbreaks, anonymous information can be accompanied by a climate of fear and mistrust. Information should ideally be provided with the source location and time for full accountability and to aid appropriate infection control (Bayham et al. 2015; Funk et al. 2009). The possibility of malevolence under such a scenario, for example by spreading rumours (intentional or unintentional), has to be taken seriously. Anonymous mobile-enabled reporting systems within these contexts could result both in poor quality data, and in the provision of a platform for increased community disharmony (Cinnamon et al. 2016).

3 The BDA Impact on Infectious Diseases Ethics

The allure of these technological applications and their impact is coupled with as yet unanswered ethical challenges. BDA-enabled infectious diseases tools could exacerbate the tensions between individual and public good. A number of these ethical questions are not novel; however, the scale at which they can occur and their potential impact has now been multiplied, as a small number of decisions/actions can now potentially affect multiple populations at different geographical locations at the same time. At what point are individuals prepared to compromise on some rights, such as that of privacy and autonomy, if it benefits them? Under what circumstances in the field of infectious diseases can individuals give up some individual benefits in order to greatly enhance public good? And, assuming that there is a benefit to the individual directly (e.g. appropriate medical treatment through better tailored response), to what extent should society be able to operate without disclosure and transparency regarding decisions taken automatically through data access and use?

3.1 Freedom of Choice

One of the key challenges in the field of BDA in infectious diseases is connected to the difficulty for individuals to be fully aware of what happens to their data after collection, as this might be generated in one locality, aggregated in a second, processed in a third, and so on. In other words, the initial data often moves through an information value chain: from data collectors, to data aggregators, to analysers/advisors, to policy makers, to implementers. Data aggregators and analysers combine the data from multiple sources and create a new picture of individuals based on the data received (Bollier 2010). Ethical issues can arise for each segment of this collection and distribution chain, even without utilizing any BDA on the data at all, with the final actor/implementer using the data for purposes that can be very different from the initial intention of the individual that provided the data.

As more data is used at increasing scale, there are growing concerns relating to aspects of automation, which is often associated to erosion of autonomy and a reduction in personal choices (Hildebrandt and Rouvroy 2011). Machine learning tries to identify relationships within big data sets and subsequently, the context in which these identified relationships are meaningful. The difficulty in accurately predicting any human behaviours based on such data correlations, lies in the tendency to ignore (or normalize) the human design and inherent error and biases in data, measures and analyses (Zuboff 2015). Furthermore, the tendencies of automation to flatten outliers and confirm patterns, to equate correlation with causation and to make it difficult to judge assumptions (i.e. lack of transparency) are also problematic (Hildebrandt 2011). The ethical question then remains: under what circumstances, if any, are such statistical approximations and reduced personal choices acceptable? The initial use of BDA as decision support can transit into an automated decision-making process, especially if the decision support uses have proven medically effective. Perhaps this is an opportunity to investigate technical mechanisms that can introduce an element of human oversight to inform and safeguard on the ethical aspects of automation, decision support and decision-making. In particular, the freedom of choice to opt-out or disagree with both decision support and decision-making outcomes should be coupled with the ability to request a second opinion that is entirely independent of the automated mechanism (e.g. an independent specialist consultation) and without any potential implications on future healthcare provision. In this manner, individuals are guaranteed a choice within existing norms and without future penalties.

3.2 Informed Consent

The reliance on algorithms for analyses in infectious diseases can introduce a gradual reduction of general understanding of the decision-making process, as the latter becomes an indecipherable black box (Pasquale 2015). Recognizing BDA as a complex process with several contributing actors including private firms is therefore important and requires transparency at each step (Asadi Someh et al. 2016; Martin 2015). Is it then ethical, or even at all possible, to achieve a truly informed consent in everyday routine clinical practice and public health, on systemic data usage in the era of BDA? Or should the model of informed consent for BDA-enabled data usage be applied fully, only to research data and clinical trials?

In an effort to navigate through some of these complexities and in the wake of the influenza pandemic, the WHO has advised that there must be a clear distinction between the boundaries of public-health oriented research and clinical practice. The informed consent of clinical practice carries the well-defined risk of preceding evidence-based practice, while the research-based activities carry a higher risk which should be made clear during the consenting process. The implementation of BDA with the potential for reduced understanding of the process also carries the risk of obfuscating the understanding of risk during consenting. Distinguishing between clinical research and practice is important, even if it is often difficult, because of the varied ways in which they are perceived and regulated in different countries (WHO 2007). The practicality of this recommendation was severely tested in the recent Ebola outbreak, where BDA-enabled

research and clinical decision-making had to be concurrent due to the nature and timings of the outbreak (WHOERT 2014).

In the case of viral diagnostics for example, the amount and granularity of information provides not only the knowledge regarding potential drug resistance parameters by the infecting organism but also the reconstruction of infectious disease outbreaks, transforming the question of ‘who infected whom’ into ‘they infected them’, i.e. from the more general to the definitive form (Escobar-Gutierrez et al. 2012; Pak and Kasarskis 2015; Bosch et al. 2016). In this context, the informed consent should be able to articulate the potential for group-level identification of infectious diseases through BDA. Most informed consent forms used currently in clinical practice identify personal risks and ignore group-level risks, perhaps as the latter are more challenging to identify and quantify.

3.3 Profiling and Justice

In order to create actionable and interpretable recommendations, aggregated population data are used to stratify the population into smaller groups to support realistic public health interventions. This method is called profiling and can occur by classifying individuals into groups, based on any given characteristic, including race, ethnic group, gender, and socio-economic status (LaBeaud et al. 2008; Newell and Marabelli 2015). However, the patient profiling in the field of infectious diseases is not always as straightforward. An important ethical aspect is the transition of an infected patient, from a vector of a disease to a potential victim, to a potential transmission node. Although technically, the link between these states can now be proven unequivocally and at increasing scale through modern diagnostics, there is a paucity on social science and ethics research exploring how these patients understand and react to these transitions—and in relation to therapeutic interventions (established or experimental). How can this status transition be ethically accounted for in the BDA era? A consent form for infectious disease research should reflect this transition and explain the context-dependent nature of the terms that are used in research. BDA should be able to accommodate the complexity arising from the possibility for patients to appear with a perceived duality of status (infecting and infected) at the same time, both at the individual and the population level.

Furthermore, the algorithms that enable profiling can ignore outliers (e.g. by reverting to mean) and provide the basis for (intentional or unintentional) discrimination among individuals or groups by downstream policy makers and implementers. As profiling (or at least some approximation) is inevitable—it is after all a very practical approach in order to rationalize the vast amounts of information into actionable information—how can individuals and group identities and differentiators be protected? Can there be safeguards in place to protect the predictable becoming exploitable? In order to achieve fairness and transparency, individuals should be made aware that decisions were taken by including profiling algorithms. There should be provision for meaningful information on the BDA approaches used and the assurance that the individual view point can be expressed, that a challenge of the decision-making through profiling will not impact future healthcare provision and that an alternative human intervention can be provided.

3.4 Surveillance and Behaviour

If effective infectious disease control policies are to be implemented in a given population, the surveillance of outbreaks needs to be coupled through BDA with modelling that predicts an individual's or a group's behaviour. Healthcare organizations can have the technological ability for the first time to continuously observe and monitor behaviours through mobile phone apps or wearable devices, offering personalized services and advice, which may imply that these individuals are no longer exposed to all options and choices that would normally be available. Is the pre-defined reduction of free choice acceptable when coupled by more effective (in terms of treatment outcome) and efficient healthcare options? This can only be decided on a case by case basis, where the reduction of free choice is coupled by specific measure to safeguard fundamental rights and the interests of the individual.

There is a substantial body of work relating to the factors that can influence population behaviour in the field of infectious diseases, using specific language to influence emotions leveraging anticipated behaviours, such as higher uptake of vaccination or reduction of contact in public spaces (Bayham et al. 2015; Chapman and Coups 2006)—and how altered social behaviour can then make a real difference in the transmission of an infectious disease within a given population (Kucharski et al. 2014; Funk et al. 2009). The impact of BDA is only expected to provide more and better calibrated opportunities to influence behaviour. Is it then ethical to control or influence an individual's or group's behaviour using information derived from their healthcare-related big data? The response to this question would inevitably be context specific and would relate to the relative balance between substantial public interest and potential harm. The proportionate nature of the pursued aim should be made transparent—even if that includes a high level of uncertainty, such as in the case of the recent Zika virus outbreaks in South America.

Additionally, the need for quick actionable advice close to real-time, (e.g. during international sports events; see Schenkel et al. 2006; Franke et al. 2006), regular cultural events, (e.g. during the Hajj; see Memish 2009) or routine operations (e.g. hospital emergency departments admissions; see Muscatello et al. 2005; Heffernan et al. 2004) complicate things further. Complexities that involve real-time predictive modelling are being addressed for unexpected and difficult to predict patterns, (e.g. climatic conditions; see Sultan 2005; Keller et al. 2009) or one-off and high-impact events (e.g. potential bioterrorist acts; see Kortepeter et al. 2000; Buehler et al. 2003). Recently, the effects of novel computing techniques and implementations have started to emerge. Algorithms can provide automated decision support as part of clinical workflow at the time and location of the decision-making, without requiring clinician initiative, and hence leading to cost and service improvements (Darrell et al. 2015). For example, as part of implementing hospital-wide analytical decision-making tools, patients presenting in emergency rooms are entered into an electronic health records system, where an algorithm can evaluate their suitability for seasonal flu vaccinations. This, in turn, prompts clinical staff to offer it, leading to a greater uptake of the vaccine and downstream savings (Venkat et al. 2010). The possibility of implementing real-time, responsive and adaptive calculations has great potential and is therefore very tempting. However, because of its direct impact on the provision of care and its resourcing in sometimes unpredictable ways, as well as its automated nature, it presents

ethicists and regulatory bodies with challenges. These centre on the potential pressures of reduced autonomy and choice, the difficulties of obtaining consent or potential for automated ‘lack of consent’ policies and even, as a worst-case scenario, the potential for unintentional harm.

4 Considerations and Recommendations

BDA has the potential to act as a catalyst and transformative agent in healthcare introducing a new era of data utilization, particularly in the field of infectious diseases. However, the ethical implications are not fully explored and if not addressed might become limiting factors preventing BDA from reaching its full potential. There are three further aspects that will be considered in this section, relating to the future developments of the BDA applications in the field of infectious diseases.

Transdisciplinary View of BDA Ethics Impact A large number of studies on BDA in healthcare reflect the uni-professional legacy of medical training and funding, yet there is a call for more collaboration between medical, life and computational sciences—often termed as ‘convergent’ or ‘inter-professional’ approach (Sharp et al. 2016; Nelson and Stagers, 2017). Such ‘convergence’ has obvious potential for new discoveries. It also requires the integration of historically distinct disciplines, technologies and ethical viewpoints into a new unified whole. The further development of BDA ethics in infectious disease will need to include an understanding of different users’ needs and capabilities, of the technologies and of the problems investigated. BDA can accommodate and perhaps be even more effective if studies conducted were based on transdisciplinary and participatory consultation designs, especially because there are so many questions yet to be defined and researched.

For example, one could conduct studies where patients would have the right to access their data, see their diagnosis/prognosis and have the ability to add their response to it (therefore both meant to collect data and intervene at the same time). These patient responses can then be used as validation inputs for BDA-based patient profiling, with the patient once again having the opportunity to view this information, applying a consistent, transparent approach throughout the patient contact points. A number of such initiatives have already taken place at different areas of the developed world. However, the usefulness of this approach is still hard to quantify as the outcome metrics used were implemented within short time frames. Additionally, there is a lack of rigorous empirical testing that can separate the effect of record access from other existing disease management programs (de Lusignan et al. 2014; Jilka et al. 2015).

However, it should be noted that the traditional metrics used for the impact of infectious diseases outbreaks still measure and report to a large extent the immediate medical impact of outbreaks—not the wider socio-economic implications. This is done for a number of reasons, such as the available infrastructure, the availability of skilled labour and the existing reporting structures, which are not able to measure the wider impact of infectious diseases outbreaks. (Heyman et al. 2015; Kruk et al. 2015) As such, any improvements to traditional approaches—through BDA or other means—do carry a potential multiplier effect in terms of their impact, where a moderate improvement in clinical outcome is still considerable within a wider socioeconomic context.

The information provided through this transdisciplinary view can include, inform and improve the ethical arguments made in relation to the public good.

Ethical Considerations Are a Competitive Advantage In the case of established infectious disease operations, the speed of technological development and the incorporation of new technologies, such as mobile phones and wearables, have outpaced the adaptability of current frameworks. For example, there is still a very slow uptake within healthcare of mobile phone-enabled appointment systems, even though the technology has existed for over a decade. Especially in the heavily regulated sector of healthcare provision, the speed of development and lack of flexibility demonstrate the need for further work in the field. One example of this type of discourse is the creation and world-wide distribution of thousands of healthcare-related applications. The majority of those are created by individuals or private providers, not institutions, and very few are regulated or if they are, they might conform to local regulations that may conflict within different regions globally. The availability and uptake of applications where individuals may voluntarily input part of their medical record has not been highlighted widely as a potential ethical minefield, even though it presents challenges not only in terms of provision of erroneous information, but also in terms of downstream personal information usage. Mobile phone diagnostics developed by healthcare providers and validated through the implementation of international standards are also present into this highly competitive melee. The FDA does not regulate applications that—in case of malfunction—do not pose a health threat to others. As such, thousands of health-related, data-rich applications remain without any regulation (Husain and Spence 2015). In light of this regulatory gap, addressing the rounded consideration of ethics can serve an acute competitive advantage for the application developer (Gibbs et al. 2016) and as a pre-condition for wider adoption into the healthcare provision system.

Open Data Policy and Education The open data movement involves governments and large public and private organizations making many of their datasets publicly available, preferably in structured machine readable formats. The underlying drive for this concept is that individual citizens, private sector and non-governmental organizations will, on a self-service basis, access and exploit these resources. However, the description of each individual is contextually dependent and informationally inexhaustible: i.e. there is no end to the potential use of one's data once created; data can be generalized, aggregated, defined, refined, repurposed, profiled and so on. The understanding of digital risks and ethical implications is lacking and should become a substantial part of open data or of data collection practices and of digital literacy training courses. As the information for an individual person is aggregated multiple times with information for many other millions of individuals and parameters, and information is extracted at multiple levels, understanding the pathway and processing of data becomes increasingly difficult for individuals to attain. In much the same manner, the attribution of accountability in the case of unintended harm through the same means becomes difficult to ascertain. As such, the recommended approach would be that of (ideally) introducing appropriate permissions for information generators and/or custodians at the point of data release and pursue transparency thereof, so that the information flow can be mapped, followed and its use audited if necessary or possible. Where knowledge dissemination is the only realistic potential benefit, e.g. following

infectious disease surveillance audits, researchers' ethical obligations need to include the understanding and description of potential collective risks in addition to the often mentioned individual identification risks.

In summary, the BDA implementation in the field of infectious diseases seems as an inevitable technological development. However, the long-term, widespread acceptance of this addition to the clinical decision-making process and adoption by patients, clinicians and the society needs to take into account the ethical aspects that are augmented or created through BDA. The need for transdisciplinary approaches to address urgent and/or complex situations should be inclusive of a transdisciplinary view of ethical needs as well. The inclusion of patients into the participatory study design and medical record access has allayed fears so far on the potential unethical use of their data through BDA. However, there are as yet no consensus metrics against which such opinions can be recorded, measured and addressed consistently. The identification of potential ethical risks at an individual and group level through BDA in infectious diseases should be a common feature, including the recognition of the contextual nature of potential ethical risks. The above are mainly viewed through the lens of the public healthcare provision; however, there are equally applicable in the case of private healthcare providers that use BDA. The latter have made great strides to develop data safety features as commercially exploitable competitive advantage and is perhaps equally relevant that the inclusion of ethical considerations can become an additional unique selling point.

5 Conclusions

The infectious diseases field's obvious mission is to provide the best possible care for infected individuals in the most efficient way, reduce the risk of outbreaks occurring and/or control existing ones. The less obvious mission is to develop preparedness for future outbreaks. The latter can be achieved through large -scale modelling and BDA utilization (Rocha and Masuda 2016). BDA in infectious diseases has not been designed as an inclusive tool, rather as a faster enabler of existing tools. The increasing availability of information through many non-traditional channels requires increased inclusivity and transparency along the needs for privacy and security. A significant number of studies have been undertaken in order to solve issues of privacy and security from a technological point of view (Safavi 2014; Camaraa 2015). That said more studies need to take place and define the ethical implication of BDA in infectious diseases in terms of (i) loss of individual autonomy and erosion of freedom of choice in response to population-level benefits and the (ii) inclusion of ethical design in the creation of BDA-enabled application aiding decision-making in healthcare provision. The success of these actions would be the wide understanding for the contextual nature of the BDA-aided interpretation and the relative risk-/benefit-based decision during infectious disease outbreaks, not just by clinicians but by the wider society affected by those outbreaks.

It is apparent that despite the positive impacts and advantages of BDA in the field of infectious diseases, the overall consequences for individuals, groups, healthcare providers and society as a whole remain poorly understood and are an under-researched topic. Importantly, an improved awareness and understanding of the ethical issues and

consequences is required to avoid the emergence of a potential negative feedback loop, where misunderstandings and lack of transparency can lead to social rejection, distorted policy and to a lack of acceptance of new technologies—limiting their potential impact for individual and societal well-being.

The surveillance potential, accuracy and immediacy of mobile-enabled technologies need to be better understood, in terms of healthcare and social impacts, including the ways in which this technology might be used for ‘social sorting’, such as the categorization of people according to risk, which can have serious real-world consequences (Lyon 2003). Irrespective of emergencies during infectious disease outbreaks, where perhaps the established state or societal norms may be temporarily suspended, the everyday potential benefits of mobile-enabled data collections need to be better balanced against their potential threats to health, freedom, non-discrimination and privacy.

One should emphasize at this point that what is now considered ‘big data’ might be only small data in decades to come. As such, the ethical pressures are expected to intensify further. The current work has highlighted the impact of BDA on the ethics of infectious diseases, as these might present themselves through established and novel perspectives. As the application of BDA in this field is expanding, further work will be necessary to define and address the ethical questions that will arise, as well as to implement consistent transparency to cultivate public trust in these evolving hyper-complex situations.

Acknowledgements This publication presents independent research supported in part by the Health Innovation Challenge Fund T5-344 (Infection response through virus genomics - ICONIC), a parallel funding partnership between the Department of Health and Wellcome Trust. The views expressed in this publication are those of the author(s) and not necessarily those of their funders and employees.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Asadi Someh, I., et al. (2016). Ethical implications of big data analytics. *Research-in-Progress Papers*, 24. http://aisel.aisnet.org/ecis2016_rip/24
- Bayham, J., Kuminoff, N. V., Gunn, Q., & Fenichel, E. P. (2015). Measured voluntary avoidance behaviour during the 2009 A/H1N1 epidemic. *Proceedings of the Royal Society B*, 282(1818), 20150814. doi:10.1098/rspb.2015.0814.
- Blais, C. M., & White, J. L. (2015). Bioethics in practice—a quarterly column about medical ethics: Ebola and medical ethics—ethical challenges in the management of contagious infectious diseases. *The Ochsner Journal*, 15(1), 5–7.
- Bollier, D. (2010). *The promise and peril of big data*. Washington: The Aspen Institute.
- Bosch, T., et al. (2016). Next-generation sequencing confirms presumed nosocomial transmission of livestock-associated methicillin-resistant *Staphylococcus aureus* in the Netherlands. *Applied and Environmental Microbiology*, 82(14), 4081–4089.
- Buehler, J. W., Berkelman, R. L., Hartley, D. M., & Peters, C. J. (2003). Syndromic surveillance and bioterrorism-related epidemics. *Emerging Infectious Diseases*, 9, 1197–1204.
- Camaraa, C. L. (2015). Security and privacy issues in implantable wearable device: a comprehensive survey. *Journal of Biomedical Informatics*, 271–289.

- Cecaj, A., Mamei, M., & Zambonelli, F. (2016). Re-identification and information fusion between anonymized CDR and social network data. *Journal of Ambient Intelligence and Humanized Computing*, 7(1), 83–96.
- Center for Global Development. (2014). *Delivering on the data revolution in sub-Saharan Africa. Center for global development and the African Population and Health Research Center*. DC: Washington.
- Chana, M. (2012). Current status and future challenges. *Artificial Intelligence in Medicine*, 56, 137–156.
- Chapman, G. B., & Coups, E. J. (2006). Emotions and preventive health behavior: worry, regret, and influenza vaccination. *Health Psychology*, 25(1), 82–90.
- Cinnamon, J., Jones, S. K., & Adger, W. N. (2016). Evidence and future potential of mobile phone data for disease disaster management. *Geoforum*, 75, 253–264.
- Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, 26(1), 77–90. doi:10.1111/isj.12088.
- Darrell, A., et al. (2014). The benefits of big data analytics in the healthcare sector: what are they and who benefits? In Wang, B., Li, R. and Perixxo, W. (eds.), *Big data analytics in bioinformatics and healthcare* (Chapter 18, pp.406-439) Hershey:IGI Global.
- Dell, N.L., et al. (2011). Towards a point-of-care diagnostic system: automated analysis of immunoassay test data on a cell phone. Proceedings of the 5th ACM workshop on networked systems for developing regions (NSDR'11). 28th of June, New York, (pp. 3-8). doi: 10.1145/1999927.1999931.
- Deville, P., et al. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45), 15888–15893.
- Doshi, P. (2009). Calibrated response to emerging infections. *BMJ*, 339, b3471.
- Duhigg, C. (2012). How Companies Learn Your Secrets? New York Times, 16 Feb 2012, (Available at http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0).
- Escobar-Gutierrez, A., et al. (2012). Identification of hepatitis C virus transmission using a next-generation sequencing approach. *Journal of Clinical Microbiology*, 50(4), 1461–1463.
- Floridi, L. (2012). Big data and their epistemological challenge. *Philosophy and Technology*, 24(4), 435–437.
- Fonkwo, P. N. (2008). Pricing infectious disease. The economic and health implications of infectious diseases. *EMBO Reports*, 9, S13–S17.
- Franke, F., Coulon, L., Renaudat, C., Euillot, B., Kessalis, N., & Malfait, P. (2006). Epidemiologic surveillance system implemented in the Hautes-Alpes District, France, during the Winter Olympic Games, Torino 2006. *Euro Surveillance*, 11, 239–242.
- Funk, S., Gilad, E., Watkins, C., & Jansen, V. A. A. (2009). The spread of awareness and its impact on epidemic outbreaks. *Proc. Nat. Aca. Sci.*, 106(16), 6872–6877.
- Gibbs, J., Gkatzidou, V., Tickle, L., et al. (2016). ‘Can you recommend any good STI apps?’ A review of content, accuracy and comprehensiveness of current mobile medical applications for STIs and related genital infections. *Sexually Transmitted Infections*, 0, 1–7. doi:10.1136/sextrans-2016-052690.
- Gilbert, G. L. (2012). Electronic surveillance for communicable disease prevention and control: health protection or a threat to privacy and autonomy? In C. Enemark & M. J. Selgelid (Eds.), *Ethics and security aspects of infectious disease control: interdisciplinary perspectives* (pp. 127–144). London: Routledge.
- Gillon, R. (1994). Medical ethics: four principles plus attention to scope. *BMJ*, 309, 184.
- GSMA. (2013). *Towards a code of conduct: guidelines for the use of SMS in natural disasters*. London: GSM Association.
- GSMA. (2014). *GSMA guidelines on the protection of privacy in the use of mobile phone data for responding to the Ebola outbreak*. London: Groupe Speciale Mobile Association.
- Heffeman, R., Mostashari, F., Das, D., Karpati, A., Kulldorff, M., & Weiss, D. (2004). Syndromic surveillance in public health practice, New York city. *EID*, 10, 858–864.
- Heyman, D. L., Chen, L., Takemi, K., et al. (2015). Global health security: the wider lessons from the west African Ebola virus disease epidemic. *Lancet*, 385(9980), 1884–1901.
- Heymann, D. L., & Rodier, G. R. (2004). SARS: a global response to an international threat. *The Brown Journal of World Affairs*, 10(2), 185–197.
- Heymann, D. L., Rodier, G. R., & The WHO Operational Support Team to the Global Outbreak Alert and Response Network. (2001). Hot spots in a wired world: WHO surveillance of emerging and re-emerging infectious diseases. *The Lancet Infectious Diseases*, 1, 345–353.
- Hildebrandt, M. (2011). The rule of law in cyberspace? Inaugural lecture mireille hildebrandt chair of smart environments, data protection and the rule of law Institute of Computing and Information Sciences (iCIS) Radboud University Nijmegen. Available at: <http://labs.sogeti.com/the-rule-of-law-in-cyberspace-by-mireille-hildebrandt/>.
- Hildebrandt, M., & Rouvroy, A. (Eds.). (2011). *Law, human agency and autonomic computing: the philosophy of law meets the philosophy of technology*. New York: Routledge.

- Hoffman, J. A., Cunningham, J. R., Suleh, A. J., et al. (2010). Mobile direct observation treatment for tuberculosis patients: a technical feasibility pilot using mobile phones in Nairobi, Kenya. *American Journal of Preventive Medicine*, 39, 78–80.
- Husain, I., & Spence, D. (2015). Can healthy people gain from health apps? *BMJ*, 350(1887), 16–17.
- Jacobsen, K. H., Aguirre, A. A., Bailey, C. L., et al. (2016). Lessons from the Ebola outbreak: action items for emerging infectious disease preparedness and response. *EcoHealth*, 13, 200–212. doi:10.1007/s10393-016-1100-5.
- Jetten, L. and Sharon, S. (2016). Selected issues concerning the ethical use of big data health analytics, 72 Wash. & Lee L. Rev. Online 486, <http://scholarlycommons.law.wlu.edu/wlulr-online/vol72/iss3/2>.
- Jilka, S. R., Callahan, R., Sevdalis, N., et al. (2015). “Nothing About Me Without Me”: an interpretative review of patient accessible electronic health records. Eysenbach G, ed. *Journal of Medical Internet Research*, 17(6), e161.
- Kaplan, W. A. (2006). Can the ubiquitous power of mobile phones be used to improve health outcomes in developing countries? *Globalization and Health*, 2, 9–10. doi:10.1186/1744-8603-2-9.
- Keller, M., Blench, M., Tolentino, H., et al. (2009). Use of unstructured event-based reports for global infectious disease surveillance. *Emerging Infectious Diseases*, 15, 689–695.
- Khan, K., Arino, J., Hu, W., et al. (2009). Spread of a novel influenza A (H1N1) virus via global airline transportation. *The New England Journal of Medicine*, 361, 212–214.
- Kortepeter, M. G., Pavlin, J. A., Gaydos, J. C., et al. (2000). Surveillance at U.S. military installations for bioterrorist and emerging infectious disease threats. *Military Medicine*, 165, ii–iii.
- Kozlakidis, Z. (2016). Biobanking with big data: a need for developing ‘big data metrics’. *Biopreservation and Biobanking*, 14(5), 450–451.
- Kozlakidis, Z., Cason, R. J. S., Mant, C., & Cason, J. (2012). Human tissue biobanks: the balance between consent and the common good. *Research Ethics*, 8, 113–123.
- Kruk, M. E., Myers, M., Varpilah, S. T., & Dahn, B. T. (2015). What is a resilient health system? Lessons from Ebola. *Lancet*, 385(9980), 1910–1912.
- Kucharski, A. J., et al. (2014). The contribution of social behaviour to the transmission of influenza A in a human population. *PLoS Pathogens*, 10(6), e1004206.
- LaBeaud, A. D., Gorman, A.-M., Koonce, J., Kippes, C., McLeod, J., Lynch, J., et al. (2008). Rapid GIS-based profiling of West Nile virus transmission: defining environmental factors associated with an urban-suburban outbreak in Northeast Ohio, USA. *Geospatial Health*, 2, 215–225.
- LaValle, S., & Lesser, E. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2), 20–32.
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, 343(6176), 1203–1205.
- Lee, E. C., et al. (2016). Mind the scales: harnessing spatial big data for infectious disease surveillance and inference. *J Infect Dis*, 214(S4), S409–S413.
- de Lusignan, S., Mold, F., Sheikh, A., et al. (2014). Patients’ online access to their electronic health records and linked online services: a systematic interpretative review. *BMJ Open*, 4, e006021.
- Lyon, D. (Ed.). (2003). *Surveillance as social sorting: privacy, risk, and digital discrimination*. London: Routledge.
- Markus, M. L. (2015). New games, new rules, new scoreboards: the potential consequences of big data. *Journal of Information Technology*, 30(1), Nature Publishing Group), 58–59. doi:10.1057/jit.2014.28.
- Martin, K. E. (2015). Ethical issues in the big data industry. *MIS Quarterly Executive*, 14(2), 67–85 1540–1960.
- Mayer-Schönberger, V., and Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt: Boston, Massachusetts.
- McCarthy, M. (2016). Slow response contributed to scale of west African Ebola epidemic, CDC concludes. *BMJ*, 354, i3814.
- Memish, Z. A., McNabb, S. J. N., Mahoney, F., the Jeddah Hajj Consultancy Group, et al. (2009). Establishment of public health security in Saudi Arabia for the 2009 Hajj in response to pandemic influenza A H1N1. *Lancet*, 374, 1786–1791.
- Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, 22, 303–341.
- Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA*, 309(13), 1351–1352. doi:10.1001/jama.2013.393.
- Muscattello, D. J., Churches, T., Kaldor, J., et al. (2005). An automated, broad-based, near real-time public health surveillance system using presentations to hospital emergency departments in New South Wales, Australia. *BMC Public Health*, 5, 141.

- Nelson, R. and Stagers, N. (2017). *Health Informatics: an Interprofessional approach*. (2nd ed.), Elsevier.
- Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision making: a call for action on the long-term societal effects of ‘datification’. *The Journal of Strategic Information Systems*, 24(1), 1–12. doi:10.1016/j.jsis.2015.02.001.
- Page, K. (2012). The four principles: can they be measured and do they predict ethical decision making? *BMC Medical Ethics*, 13, 10.
- Page, T. (2015a). A forecast of the adoption of wearable technology. *International Journal of Technology Diffusion*, 6, 12–29.
- Page, T. (2015b). Privacy issues surrounding wearable technology. *I-Manager's Journal on Information Technology*, 4, 1–16.
- Pak, T. R., & Kasarskis, A. (2015). How next-generation sequencing and multiscale data analysis will transform infectious disease management. *Clinical Infectious Diseases*, 61(11), 1695–1702.
- Pasquale, F. (2015). *The black box society: the secret algorithms that control money and information*. Cambridge: Harvard University Press.
- Pfaff, G., Lohr, D., Santibanez, S., et al. (2010). Spotlight on measles 2010: measles outbreak among travellers returning from a mass gathering, Germany, September to October 2010. *Euro Surveillance*, 15, 5–8.
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3.
- Randrianasolo, L., et al. (2010). Sentinel surveillance system for early outbreak detection in Madagascar. *BMC Public Health*, 10, 31. doi:10.1186/1471-2458-10-31.
- Richards, N. and King, J. (2014). ‘Big Data Ethics’, Wake Forest Law Review. Available at SSRN: <http://ssrn.com/abstract=2384174>.
- Rocha, L. E. C., & Masuda, N. (2016). Individual-based approach to epidemic processes on arbitrary dynamic contact networks. *Scientific Reports*, 6, 31456. doi:10.1038/srep31456.
- Safavi, S. (2014). Conceptual privacy framework for health information on wearable devices. *PLoS One*, 9, 9–12.
- Schenkel, K., Williams, C., Eckmanns, T., et al. (2006). Enhanced surveillance of infectious diseases: the 2006 FIFA World Cup experience, Germany. *Euro Surveillance*, 11, 234–238.
- Segura Anaya, L. H., Alsadoon, A., Costadopoulos, N., et al. (2017). Ethical implications of user perceptions of wearable devices. *Science and Engineering Ethics*, 1–28. doi:10.1007/s11948-017-9872-8.
- Sharp, P., Jacks, T., & Hockfield, S. (2016). Capitalizing on convergence for health care. *Science*, 352(6293), 1522–1523. doi:10.1126/science.aag2350.
- Simonsen, L., Gog, J. R., Olson, D., & Viboud, C. (2016). Infectious disease surveillance in the big data era: towards faster and locally relevant systems. *The Journal of Infectious Diseases*, 214(4), S380–S385. doi:10.1093/infdis/jiw376.
- Smith, M., et al. (2012). Big Data Privacy Issues In Public Social Media. In: 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST 2012) (pp. 1–6).
- Story, A., Garfein, R., Hayward, A. C., et al. (2016). Monitoring therapy compliance of tuberculosis patients by using video-enabled electronic devices. *Emerging Infectious Diseases*, 22(3), 538–540.
- Stuart Ward, J. and Barker, A. (2013). Undefined by data: A survey of big data definitions. pre-print deposition, arXiv:1309.5821v1.
- Sultan, B., Labadi, K., Guegan, J. F., & Janicot, S. (2005). Climate drives the meningitis epidemics onset in west Africa. *PLoS Medicine*, 2, e6.
- Trad, M-A., Jurdak, R., Rana, R. (2015). Guiding Ebola patients to suitable health facilities: an SMS-based approach. F1000Res, 4, 43.
- Vanhems, P., Barrat, A., Cattuto, C., Pinton, J.-F., Khanafer, N., Régis, C., et al. (2013). Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS One*, 8(9), e73970.
- Venkat, A., Chan-Tompkins, N. H., Hegde, G. G., Chuirazzi, D. M., Hunter, R., & Szczesiul, J. M. (2010). Feasibility of integrating a clinical decision support tool into an existing computerized physician order entry system to increase seasonal influenza vaccination in the emergency department. *Vaccine*, 28(37), 6058–6064.
- WHO Ebola Response Team (WHOERT). (2014). Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *The New England Journal of Medicine*, 371, 1481–1495.
- Wigan, M. R., & Clarke, R. (2013). Big data’s big unintended consequences. *IEEE Computer*, 46(6), 46–53. doi:10.1109/MC.2013.195.
- World Health Organization. (2007). *Ethical Considerations in Developing a Public Health Response to Pandemic Influenza*. Geneva: Document WHO/CDS/EPR/ GIP/2007.2.

- Yang, C., Yang, J., Luob, X., & Gong, P. (2009). Use of mobile phones in an emergency reporting system for infectious disease surveillance after the Sichuan earthquake in China. *Bulletin of the World Health Organization*, 87, 619–623.
- Yeslam, A.-S. (2015). The use of data mining by private health insurance companies and customer's privacy. *Bioethics and Information Technology*, 24, 281–292.
- Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), Nature Publishing Group), 75–89. doi:10.1057/jit.2015.5.