Evidence and future potential of mobile phone data for disease disaster management

Jonathan Cinnamon, Sarah K. Jones, W. Neil Adger

Forthcoming in Geoforum

Abstract

Global health threats such as the recent Ebola and Zika virus outbreaks require rapid and robust responses to prevent, reduce and recover from disease dispersion. As part of broader big data and digital humanitarianism discourses, there is an emerging interest in data produced through mobile phone communications for enhancing the data environment in such circumstances. This paper assembles user perspectives and critically examines existing evidence and future potential of mobile phone data derived from call detail records (CDRs) and two-way short message service (SMS) platforms, for managing and responding to humanitarian disasters caused by communicable disease outbreaks. We undertake a scoping review of relevant literature and in-depth interviews with key informants to ascertain the: i) information that can be gathered from CDRs or SMS data; ii) phase(s) in the disease disaster management cycle when mobile data may be useful; iii) value added over conventional approaches to data collection and transfer; iv) barriers and enablers to use of mobile data in disaster contexts; and v) the social and ethical challenges. Based on this evidence we develop a *typology* of mobile phone data sources, types, and end-uses, and a *decision-tree* for mobile data use, designed to enable effective use of mobile data for disease disaster management. We show that mobile data holds great potential for improving the quality, quantity and timing of selected information required for disaster management, but that testing and evaluation of the benefits, constraints and limitations of mobile data use in a wider range of mobile-user and disaster contexts is needed to fully understand its utility, validity, and limitations.

Keywords: mobile phone; call detail records; SMS; disaster; disease; big data

1 Introduction

Novel sources of data coupled with new data mining techniques and data-driven logics are transforming science, business, governance and society more generally. In this age of so-called *big data*, there is potential for transformative change in the role of information in decision-making as we move from a mode of decision-making defined by data scarcity to a new era of data abundance (Hey *et al.*, 2009; Miller, 2010; Miller & Goodchild, 2015). The potential of new sources of data to address global data inequities holds particular promise in low- and middle-income countries (LMIC) where conventional sources of social, environmental, and economic data are often patchy, many years out of date, or simply non-existent (Center for Global Development, 2014; Cinnamon & Schuurman, 2013; Deville *et al.*, 2014). Kitchin (2013) describes three approaches to the way big data are produced: *directed* – in which a human operator focuses a data capturing technology on a person or place (e.g. surveillance camera, remote sensing); *automated* – in which data are passively collected via the normal operation of a system or technology (e.g. mobile phone use, Web browsing, credit card transactions); and *volunteered* –

data which are actively or passively produced by citizens, typically via user-generated platforms including social media and crowdsourcing applications. These *active* and *passive* approaches to big data production can rapidly produce new sources of data in real or near-real time, which opens up a range of opportunities for understanding diverse social phenomena, especially for 'data poor' settings characterized by inadequate data infrastructures. The potential of big data to improve disaster response and management is stimulating significant interest from researchers, government, and the humanitarian community (e.g. Fadiya *et al.*, 2014; Pu & Kitsuregawa, 2013; Shelton *et al.*, 2014; Zoomers *et al.*, 2016). This paper critically examines the existing evidence and explores the future potential for the use of actively and passively produced mobile phone data for managing humanitarian disasters caused by communicable disease outbreaks.

A disaster can be defined as a "serious disruption of the functioning of a community or a society involving widespread human, material, economic or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its own resources" (UN International Strategy for Disaster Reduction, 2009, p. 9). Although attention and resources are concentrated on responding to disasters, the *disaster management cycle* (Figure 1) is commonly conceptualized as having four overlapping and interconnected phases focusing on risk management (*prevention* and *preparedness* phases*)* and crisis management *(response* and *recovery* phases) (Alexander, 2002, p. 6). Together these phases represent the "sum total of all activities, programmes and measures which can be taken up before, during and after a disaster with the purpose to avoid a disaster, reduce its impact or recover from its losses" (Vasilescu *et al.*, 2008, p. 46).

Figure 1: Phases of the disaster management cycle

Rapid evidence-based response to events such as natural disasters, disease outbreaks, and political emergencies is essential to minimize losses and damage, and build community resilience. To decide how to distribute emergency resources in the immediate aftermath of a disaster, relief agencies need access to information about the magnitude of the event, the locations where people have been impacted, population characteristics and dynamics, and existing distribution of relief resources and infrastructure (MapAction, 2016). Data inadequacies can result in excessively slow, ineffective, and in some cases negligent disaster response and recovery as illustrated by the widely criticized response to Hurricane Katrina (Thompson *et al.*, 2006). Poor decision-making during the response and recovery phases can be amplified by inadequate knowledge about vulnerabilities during the prevention and preparedness phases of the disaster management cycle (Pu & Kitsuregawa, 2013), highlighting the importance of access to data through all phases of disaster management.

A range of data types can support disaster management. Some directed sources of big data such as remotely-sensed satellite imagery are already relatively widely used in disaster management (Voigt *et al.*, 2007). Use of other direct data sourcing technologies, notably unmanned aerial vehicles (UAV), is not yet standard practice but hold great promise for rapid access to high spatial and temporal resolution imagery from before and after natural disasters (Griffin, 2014) and for enhancing situational awareness in humanitarian emergencies (Aubrecht *et al.*, 2015). Similarly, the use of volunteered sources of data in disaster management has been widely reported in recent years as part of the field of 'digital humanitarianism' (Burns, 2014; Meier, 2015), particularly the passive harvesting of spatial information about crisis events from social media including geotagged tweets (Shelton *et al.*, 2014; Zook *et al.*, 2010), and active production of data using Web-based crowdsourcing platforms (Goodchild & Glennon, 2010; Roche *et al.*, 2013). These technologies are transforming communications during emergencies, however the representativeness, credibility, social and ethical consequences, and overall value of the data for decision-making during disasters is not well understood (Buscher *et al.*, 2014; Crawford & Finn, 2015; Shanley *et al.*, 2013).

Comparatively, actively and passively produced big data generated through mobile phone communications have received considerably less attention from within the academic and humanitarian communities and is an underused data source in disaster management (Madianou *et al.*, 2015; The Economist, 2014), despite the rapid global diffusion of mobile phones in recent years. Mobile networks now cover 50% of the globe, supplying almost 8 billion mobile phone connections, and over half of the world's population is estimated to have a mobile subscription (GSMA, 2015). Although worldwide penetration of mobile phone usage is uneven – loosely paralleling global patterns of development – access to mobile phones is one metric for which many LMIC are rapidly catching up with the rest of the world. In sub-Saharan Africa, 50% of the adult population are expected to have mobile phones by 2020, with some countries far surpassing this already, including Nigeria (89%), Senegal (84%), and Ghana (82%) where subscription rates are similar to many high-income countries, although smartphones are still relatively rare in this region (Pew Research Center, 2015). These data assume that each mobile phone owner is independent while in reality some or even many individuals may be owners of two or more mobiles, meaning coverage could be more limited than it appears. Nonetheless, there is no doubt that the number of mobile phones in use in LMIC is increasing. This technological leapfrogging to digital communication methods in LMIC has been driven by its use for a range of purposes beyond personal communications, including mobile money (digital bank accounts and financial services) which can be accessed via short message service (SMS) on basic mobile phones.

In this paper, we critically examine the current and future potential of two types of data produced via mobile phone communications: passively produced call detail records (CDRs), and active produced data through SMS. Mobile communications produce massive longitudinal datasets recorded and maintained by mobile network operators (MNO). These CDRs typically include the time a communication was made, a unique identifier for the caller, receiver's telephone number, call duration, size of data transmitted, and the geographic location of the cellular tower the call was routed through and received from for every communication (call or SMS) made by every mobile phone user (FCS, 2012). CDRs are routinely collected by MNOs to facilitate customer billing, problem diagnostics, and network planning, and they have a massive potential to illuminate the spatio-temporal dynamics of individuals and populations at a very high resolution in near real-time (Deville *et al.*, 2014). Researchers are starting to tap into these data sources to understand population characteristics (Douglass *et al.*, 2015), transportation and mobility (Doyle *et al.*, 2014), socio-spatial behaviours and interactions (Gao *et al.*, 2013; Järv *et al.*, 2014), urban spatio-temporal dynamics (Ahas *et al.*, 2015) and inferred aggregate economic activity (Scepanovic *et al.*, 2015). CDRs are thus poised to significantly advance knowledge in these areas, especially in domains that have conventionally relied on out-dated, unrepresentative, or low-resolution data sources (e.g. national census, surveys, modelled estimates).

Mobile phones can also be used to communicate and gather data directly from citizens via mobile apps and SMS. We focus on SMS, as this is available to all types of mobile phone – basic, feature, and smart – and may therefore be the most appropriate starting point for mobile-based disaster management in LMIC. SMS technology has been available since the 1990s and has been widely used as a medium for *sharing* information with the public during emergencies, however it is only recently that it has also seen significant use for actively *collecting* timely disaster information. Interactive *two-way SMS* communication and data collection is a survey method that works by sending questions to all or a targeted selection of mobile phone subscribers. The receiving database can be configured to trigger further questions depending on the user's response. Two-way SMS is a reliable, acceptable, and low-cost tool for collecting data for research purposes (Whitford *et al.*, 2012) and is a promising platform for actively engaging mobile phone users to volunteer spatio-temporal data in disaster contexts, including information on personal vulnerability and resilience, local impacts, and resource requirements, while also helping to build trust and encourage local resilience (McDonald, 2016).

These sources of "Big Mobile Data" (Ahas *et al.*, 2014, p. 5) could be particularly effective in global health emergencies, such as communicable disease outbreaks, because of the importance of tracking population movements (Zhong & Bian, 2016) and rapidly responding to needs in this context, e.g. to minimise contact between infected and uninfected populations. However, there is a lack of clear evidence on benefits, challenges and limitations of mobile data to guide its effective and appropriate use in health emergencies and disaster contexts more generally. As relatively new data sources, this gap raises the risk of information misuse if the limitations of mobile data sources are not well-understood, and of missed opportunities to improve disaster management if the advantages of these data are poorly recognised. In this paper we seek to help address this knowledge gap by assessing the opportunities, limitations, constraints, and enablers to using mobile data in disaster contexts. While there are only a small number of early exploratory examples of CDR or SMS data use in disasters (e.g. Bengtsson *et al.*, 2011; Yang *et al.*, 2009), the 2013-2015 Ebola outbreak and humanitarian crisis in West Africa stimulated significant interest and debate in academic, policy, and humanitarian circles around the use of mobile phone data in disease disaster management (e.g. The Economist, 2014; The Guardian, 2014; Vayena *et al.*, 2015; Wesolowski *et al.*, 2014a), against the backdrop of a "growing narrative that the problem in the response effort was a lack of good information technology and, more specifically, data" (McDonald, 2016, p. 3). We systematically review the growing body of literature from previous use cases, and conduct semi-structured interviews with key informants with hands-on experience of humanitarian response, to develop a knowledge base and explore the utility, value, and challenges of CDR and SMS data for responding to and managing humanitarian disasters. These empirical findings are then used to develop a typology for mobile data in disease disaster contexts and a decision-tree that could be used to assist data use decisions. This study aims to advance discussion around the use of these data, by assembling and critically engaging with the issues that should be considered by researchers, humanitarian organizations, and policy makers who wish to make use of SMS or CDR data for disaster management. Although the focus is on disease disasters, the findings are likely to have relevance for a diverse range of disaster and humanitarian contexts.

2 Mobile phone data and communicable disease disasters: Assembling the evidence

We developed a two-step methodology to assemble evidence and gather perspectives regarding mobile data use in disease disaster management. First, we followed Arksey and O'Malley's (2005) widely used scoping review approach, which is a type of systematically-conducted literature review that aims "to map rapidly key concepts underpinning a research area and the main sources and types of evidence available" (Arksey & O'Malley, 2005, p. 21). A literature search strategy was devised which focused on academic, grey literature (e.g. organizational reports and other works not distributed by commercial/academic publishers), and media sources published between 1 January 2009 and 15 May 2015 to capture the rapid expansion of evidence and debate emanating from the Ebola crisis and the small amount of relevant literature published prior to this event. The academic search databases used were Google Scholar, Science Direct and PubMed. Media sources were identified using Nexis UK's 'Major World Publications' and Google News. Google Web search engine was used to identify any grey literature including organisational reports and websites. Based on an initial scan of the literature, the research team decided to focus on extracting information relevant to five main thematic areas:

1. Information that can be gathered from CDR or SMS data
2. Phase(s) in the disease disaster management cycle when mobile data may be useful
3. Value added over previous/conventional approaches to data collection and transfer
4. Barriers and enablers to use of mobile data in disaster contexts
5. Social and ethical challenges relating to mobile data use

Search terms used were *Ebola, disease, health, emergency, disaster, big data, mobile phone, cell phone, SMS, call detail records, text message, and information needs*. Based on these search terms, a total of 48 English language pieces were included in the review. Of these, 21 were from peer-reviewed journals (empirical studies and overview pieces), 20 were media sources (traditional and new media articles), and 7 were grey and organizational literature (e.g. reports, blog posts). A majority (25) of the pieces were explicitly or primarily about the use of mobile data to combat the 2013-15 Ebola crisis.

Next, we conducted ten in-depth, semi-structured interviews with key informants from humanitarian agencies, international aid organizations, mobile data collection service providers, and mobile data focused research teams. The respondents were selected due to their direct experience of using mobile phone data in disaster or disease management contexts, or due to their experience in information management in disease-related disaster situations. The selection involved our initial professional networks, followed by snowball sampling to purposively seek highly informed and relevant international actors in this area. The interview questions were loosely structured around the same five themes used for the review, but kept open-ended to encourage participants to share all information they felt was relevant.

3 Results: Call detail records and SMS data for disease disaster management

This section presents the findings of the scoping review and interviews in thematic order.

3.1 Information that can be gathered from CDR or SMS data

CDR datasets contain information on the location and time at which a communication (call/SMS) is made along with unique identifiers for the sender and receiver. These data may be valuable for disaster management when they are used to estimate population size and density in a region or city, based on the number of phone subscribers in the coverage area of each cellular tower (Bharti *et al.*, 2015). Information on population distribution before and after a disaster can be useful when assessing exposure risk and response needs. In a health emergency, the high spatial and temporal resolution data on population numbers inferred through CDRs could assist in identifying potential pathways of disease transmission, populations and locations at risk of disease outbreak, population-level access to and need for medical and social services, and effective allocation of humanitarian aid (Bharti *et al.*, 2015).

Significant recent attention has focused primarily on the use of CDRs to assess spatial patterns of human travel and for tracking population mobility over time (Tatem, 2014; Tatem *et al.*, 2014; Wesolowski *et al.*, 2014a; Wesolowski *et al.*, 2014b). For diseases like Ebola which are spread by person to person contact, this use of CDRs can be valuable for estimating population migration from areas of high infection to other areas, and may indicate important routes of travel (Wesolowski *et al.*, 2014b). These population movements are derived by tracking the geographic location (i.e. based on the nearest cellular tower when a communication is made by the mobile phone user) of individual mobile numbers over a period of time, which provides information on the number of mobile users that remain in an area, or leave, and their direction of travel. Tatem *et al.* (2014) describe how population movements between 'source' and 'sink' (net exporter and importer of diseases) communities inferred from CDRs can be combined with disease case data to create maps of disease risk. Population location information gathered via CDRs can further be used to assess compliance with travel restrictions in disease-affected areas, or government closure orders of facilities (Oliver, 2013). When unique identifiers are included in CDR datasets, it may even be possible to assess individual-level mobility. For example, Tatem et al. (2014) describe how a 'metric of cumulative risk' can be calculated for individuals, based on the total amount of time spent in a disease endemic region.

SMS-based tools are, meanwhile, enabling humanitarian agencies to collect specific data on population needs when mobile phone users respond to questions sent via text message. The literature describes the use of SMS platforms for collecting information from the public and from health professionals, such as for reporting of potential disease symptoms or supply requirements (Asiimwe *et al.*, 2011). For example, Tracey *et al.* (2015) describe EbolaTracks, an SMS based platform for active self-monitoring of persons who recently visited an Ebola outbreak area. The person is provided with a self-monitoring kit including thermometer, and a mobile phone with SMS capability. Twice daily over the next 21 days (Ebola incubation period), health personnel send text messages to the persons involved in the programme, instructing them to report symptoms and temperature. Any abnormal reports alert a medical officer, who proceeds to follow up the potential case and initiate an isolation procedure if required.

During a health emergency, two way SMS can also work as both a triage tool, and as a crowdsourced data collection platform for collating information from the public. Several open source two-way SMS data platforms exist, including Frontline SMS (http://www.frontlinesms.com/), and RapidPro SMS (https://community.rapidpro.io/), facilitating data collection and analysis. Trad et al. (2015) describe a proposed SMS platform that would both direct suspected Ebola cases to nearby health facilities, as well as collect information from the public about disease cases. SMS is also being used to collect other vital information needed to effectively respond to health emergencies, including availability of resources in an area. In 2014 amidst the Ebola outbreak, the World Food Programme set up a monthly SMS-based survey for all phone users in Sierra Leone, Guinea and Liberia, asking people to report food shortages and prices in their area (Anderson, 2014). The RapidPro open-source SMS system was deployed in Sierra Leone to quickly gather information on Ebola prevention supplies in schools, including washing kits and thermometers, enabling rapid response to those schools suffering shortages (Mushayi, 2015).

3.2 Phases in the disease disaster management cycle when mobile data may be useful

Figure 2 presents a typology for describing the source, type and primary end-use of passively and actively produced mobile phone data during different phases in the disease disaster management cycle, based on information emerging from our analysis. Note that smartphone app and interactive voice response (IVR) data are not discussed here but are likely to have similar characteristics as two-way SMS data.

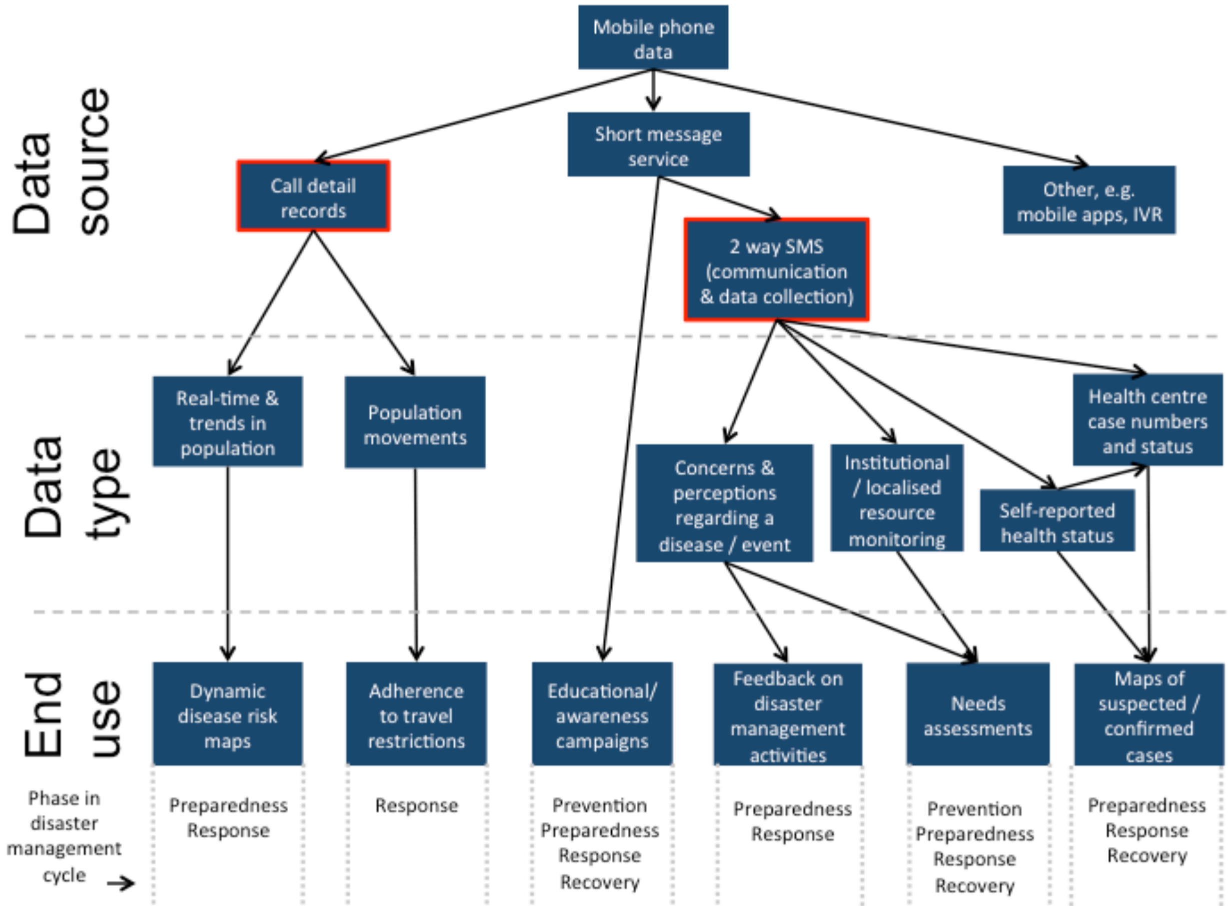

Figure 2: Typology of mobile data sources, types and uses in the disease disaster management cycle

3.3 Value added over previous/conventional approaches to data collection and transfer

The value added of mobile data over conventional sources and methods of data collection for disaster management relate to the potential for *improved data availability and accuracy*, *more rapid detection/response,* and *increased efficiency*. In many LMIC in particular, mobile data could provide information that are unobtainable, incomplete, or out of date from other sources, pointing to the possibility for these data sources to fundamentally disrupt conventional approaches to disaster decision-making in settings defined by data scarcity. Census data on population size and mobility, if they exist, do not account for short term movements or visits, the category of mobility of most interest in communicable disease disaster management (Hern, 2014). The rapid growth of mobile phones in low- and middle-income countries, and the CDRs produced through their use, provide access to population figures and mobility information not available from other sources (Wesolowski *et al.*, 2014a). CDR data are typically of high spatial and temporal resolution and broad in geographic scope, covering the operator's network area which is often the entire country, which represents a huge value added over other population datasets derived from the census or surveys (Hern, 2014). Disaster response teams often rely on data from the International Organisation for Migration (IOM) to track population movements, which are collected by

IOM personnel counting the number of people moving through official border posts. This provides a reliable source of data to humanitarian agencies - "*We are very trusting of IOM methods*" (Interview participant, NGO employee, 28 April 2015) – but these data are only available when IOM are in-country, are labour and time intensive to collect, and do not account for unofficial border crossings or within-country movements. In contrast, CDR data can be used for tracking within country population movements, are continuously collected, can be made available very rapidly once the right agreements are in place, and have little to no on-going data collection costs, since the data are generated automatically (Lydersen, 2014; Tatem *et al.*, 2009), which opens up new opportunities for information generation in an emergency. With respect to the use of CDR data for understanding the spatio-temporal dynamics of populations, one participant noted "*[w]e're facing the mapping part on better more rapidly reported data*" (Interview participant, academic researcher, 10 November 2015). In short, the key value added of CDR data over census or survey approaches is the potential to access current and comprehensive evidence on population size, density, and dynamics, information that is fundamentally necessary for managing any humanitarian emergency or disease-related disaster but which is often unavailable or insufficient in many settings, and especially so in most LMIC.

Rapid decision-making based on evidence is difficult without suitable data sources and adequate quality outputs. Mushayi (2015) describes how an SMS system rapidly deployed in the Ebola response in Sierra Leone enabled discussions on resource decisions to be based on evidence rather than anecdotes or guesswork. The availability of SMS data could stimulate an acceleration of knowledge and the exchange of time sensitive information such as resource needs or new disease cases during disease emergencies, which could have a positive influence on efficiency in terms of costs and resource usage (Revere *et al.*, 2014). As one interview participant noted, "*the speed and ease with which you could get some data at the country level with [SMS] would be hard to do in any other way*" (mobile data platform service provider, 11 August 2015). In the aftermath of the 2008 earthquake in the Sichuan area of China, an SMS based infectious disease reporting system was developed and tested within 3 days, and was in use one week later across the region (Yang *et al.*, 2009). Speaking about the value added of SMS over face-to-face data collection approaches, a participant who provided mobile data services highlighted the broad geographic reach and potential representativeness that could be achieved with SMS methods: "*people weren't going to be satisfied with a few focus group discussions around key parts of the response, they wanted more "robust" data. So to get that we needed data at some kind of scale, and mobile phones were the only way we could have done that*" (11 August 2015). Many data collection activities in low-income countries typically use paper-based records, such as those kept by community health centres. For this, paper-based data need to be made physically accessible to the data collector, which can be a huge barrier to time sensitive decision-making in an emergency situation. Interview participants with experience in this type of data collection stressed the high value of SMS based data transfer for substantially increasing efficiencies, facilitating multi-level communication, reducing data transfer errors (even though human errors in data entry may be the same), and facilitating standardisation of data records. "*I think the first thing is that your operational complexity goes down by an order of magnitude as soon as you use a mobile data collection as compared to a face to face data collection*" (Interview participant, mobile data platform service provider, 26 August 2015). In short, SMS based methods can enable much faster and more widespread communication and collection of time sensitive information during emergency situations. This could accelerate the identification and isolation of potential communicable disease cases, reduce the burden on health care personnel through patient self triage, and encourage more appropriate allocation of supplies, resources and health personnel compared to conventional methods.

3.4 Barriers and enablers to use of mobile data in disaster contexts

Researchers and organizations should be aware of a range of issues that may prevent or reduce the use of CDR and SMS data in disease related disaster management. These relate primarily to *data access, survey design limitations, data processing, difficultly of collecting qualitative or contextual information, problems with obtaining representative population samples* and *uncertain data quality and validity.*

Perhaps the most significant barrier to the use of CDR data is the fundamental and highly complex challenge of data access. Government regulators have shown limited political will or interest in developing protocols and regulation for accessing CDRs (Shacklett, 2015; The Economist, 2014). Gaining access to CDR data is especially challenging in the context of tracking international migration – which is of significant interest in global health crises - because of the need to liaise with multiple mobile operators and national regulators, and users switching SIM cards as they cross borders. CDR access regulations are not internationally standardized (de Montjoye *et al.*, 2014); as one participant described, "*the regulations are changing month by month, week by week sometimes, in terms of what needs to be done to meet the national regulations on how these kind of data can be processed and accessed*" (interview participant, academic researcher, 10 November 2015). There is a lack of clarity around data protection laws within countries, and laws vary greatly between countries; during the height of the Ebola outbreak centred on Guinea, Sierra Leone, and Liberia, MNOs in the first two countries released CDRs while more restrictive data protection prevented their use in Liberia (McDonald, 2016). Even if data protection regulations enable access to CDRs, MNOs may be unwilling to release these data to third parties (Lydersen, 2014; WorldPop, 2014) because of remaining concerns about customer confidentiality as well as the disclosure of proprietary information regarding their database and system design (Wesolowski *et al.*, 2014a). CDR data access challenges were voiced by several interview participants: "*Probably what most of us spend a huge amount of our time doing is talking to operators and health and national regulators, as well as operators at the group level and the country level. Data access is probably the biggest challenge at the moment.*" (Interview participant, academic researcher, 10 November 2015). These challenges include getting access to the geographic information (cell tower location) essential for spatio-temporal analysis and mapping of CDR data, which is not necessarily included as standard practice by mobile operator databases: "*We've tried to work with MNOs before on geographic info and have never had any success… We always get CDRs but I have never seen a CDR with geographic info.*" (Interview participant, mobile data platform service provider, 26 August 2015). Buckee *et al.* (2013) describe how operators release CDRs in different formats, meaning tools have to be modified for each release. The same authors note that population mobility patterns are culture-specific, so CDR based population models and tools developed in Western settings may not apply to the mobility patterns of LMIC, such as those of seasonal pastoralists, refugees and migrant workers.

This points to one of the most significant potential enablers to mobile data use: improved protocols and guidelines for data access, processing, and sharing. The Groupe Speciale Mobile Association (GSMA) has developed guidelines for the appropriate use of CDR and SMS data in emergency situations (GSMA, 2013, 2014), which outline amongst other things, best practices in terms of stakeholder agreements, access protocols, and data security measures. These initial guidelines are helping to clarify the benefits and risks of mobile data for all stakeholders, and further refinement and detail in future iterations should help to accelerate their use in emergency contexts. The mounting media attention given to CDR and its value in disaster contexts may also be helping to overcome data access barriers. For example, discussions on the social value of CDRs during emergency situations may encourage MNOs to share data and collaborate with stakeholders as part of corporate social responsibility programs. As Vayena *et al.* (2015) describe, health is a common good, so making CDRs available for responding to and managing health emergencies may be more acceptable to the public, governments, and MNOs than for other uses, however the issue of purpose limitation is a concern. Stakeholders are likely to be more willing to release CDRs in disease and emergency management, however they may have reservations if they feel unable to control any subsequent uses of the data.

Data access challenges are further being reduced by the growing market for mobile data service providers that facilitate data access for the NGOs and humanitarian organizations that have the funds to secure their services. There are now a number of companies that have agreements in place with MNOs in various countries and can act as intermediaries between the network and those wishing to access it to obtain CDR records or for SMS based communications. Organizations that wish to access mobile networks for these purposes can do so rapidly without the need to set up bespoke agreements with operators, and the intermediaries can also provide services to assist the organization with setting up effective surveys and messaging. However, even once access to CDRs has been agreed, there are still significant barriers to use due to the time and cost associated with data cleaning and preparation (Jacobsen, 2014). Analysing and managing massive and complex CDR datasets requires advanced technologies and data science expertise (Asokan & Asokan, 2015; Young, 2015). Undertaking sophisticated analyses without the requisite analytical skills or domain knowledge can lead to spurious conclusions (Macharia, 2015) and deter agencies without in-house expertise from attempting to use CDRs. Even if CDRs are high quality and skills and technologies are available, decision-making models and outputs are only as good as the other datasets used, such as disease case locations which are often patchy in many settings (Bengtsson *et al.*, 2015).

Concern over data quality and representativeness is another significant barrier to the effective use of CDRs. Although penetration has rapidly increased in many LMIC, some countries have comparatively lower rates of ownership including those affected by Ebola in West Africa (Lydersen, 2014). With phone ownership, there tends to be a bias towards the more wealthy and mobile, meaning the results may not necessarily reflect the wider population's movements (Tatem *et al.*, 2014), and may in fact be an overestimate (see also Xu *et al.*, 2016; Zhao *et al.*, 2016). Also, rural areas may not be covered or may have very large tower catchment areas, meaning the spatial resolution will be poor (Lydersen, 2014; Shacklett, 2015). However, one interview participant expressed the view that coverage should not be viewed as a major constraint with the ease at which rapid assembly mobile towers can be set up, which were successfully used in Sierra Leone by UNICEF to improve network coverage. For those communicable diseases spread by person to person contact such as Ebola, the lack of representativeness outlined above is perhaps less problematic as these diseases tend to cluster in dense populations (where penetration is typically higher), due to the mode of transmission (Buckee *et al.*, 2013). Further, it is expected that mobile ownership and rural penetration will continue to rise significantly in the coming years, meaning demographic and geographic representativeness should steadily improve (Wesolowski *et al.*, 2014a).

SIM card practices in LMIC are an added barrier to data quality (Wesolowski *et al.*, 2014b). It is not uncommon in LMIC for individuals to have several SIM cards and use them at different times of day or location, complicating the tracking of multiple SIM users over time and between locations. Phone-sharing practices between friends and family members adds another level of complexity. As one interview participant remarked, "[*w*]*e know that all we're tracking is SIM cards and we don't really know whose they are or how representative they are of the population and then it's another step beyond that to link them to disease and infections moving around*" (Interview participant, academic researcher, 10 November 2015).

In comparison to CDR, use of SMS data is currently more common in the humanitarian sector, although there remain a number of barriers to their use. Similar to CDRs, access is a challenge when trying to set up SMS data collection platforms. Telecommunication infrastructure requirements for SMS based data collection can prevent access to these data in time sensitive emergency situations: "*I* *think the issue with the emergency situations is that it is very difficult to deal with telcos [telephone companies] to set up the infrastructure, and so in an emergency situation you need to move quick, and basically if you don't have it set up already, you probably have no hope of getting it set up*" (Interview participant, mobile data platform service provider, 26 August 2015). Here, infrastructure refers to a centralised data centre set up to receive or send data through mobile line connections, and the software that enables this information to be managed. While it is not essential to collaborate with MNOs to set up SMS data collection systems, this participant explained that doing so has many advantages, not the least the ability to access the phone numbers of potential SMS survey participants in the region of interest. Indeed, humanitarian use of the networks to broadcast information and gather data without usage costs requires agreements between the mobile operator, governments and humanitarian agencies, which may be time consuming (O'Donovan & Bersin, 2015).

SMS are limited to 160 characters per message and relying on simple or truncated communications can have significant limitations for both communicating and gathering vital and time sensitive health information (Revere *et al.*, 2014). This points to a wider issue in that SMS based data collection requires a relatively simple survey design and this may pose a challenge for users accustomed to more sophisticated survey designs. "*People are really in a mindset of, like, I need to ask someone 200 questions or it's not worth it. And you know, 200 questions is pretty hard to do over the phone.*" (Interview participant, mobile data platform service provider, 26 August 2015). SMS is best for collecting basic closed-answer questions, for instance using a Likert scale or binary yes/no structure. As one participant described "*open questions are really difficult to do, we have done them in the past, and the quality really, really ranges … from absolutely useless to really in depth*" (interview participant, mobile data platform provider, 11 August 2015). In depth qualitative information is thus challenging to acquire via SMS, and can lead to the collector receiving an incomplete picture or failing to interact with a participant in a sufficiently sensitive manner. Subtle information provided by a respondent's tone of voice or body language is not captured in an SMS survey, while this can be systematically recorded during face-to-face interactions with individuals, or questions that allow open-ended responses. In addition several participants described how the presence of multiple spoken languages as well as low rates of literacy and numeracy in some settings can inhibit the potential of SMS data collection and should be carefully considered before conducting SMS-based surveys.

The increasing recognition of crowdsourcing and citizen science in society could be seen as an enabler of SMS based data collection (Gilpin, 2014), however this also presents a major barrier to the effective use of these data for decision-making. Relying on often-anonymous members of the public or health workers to provide reports on personal disease symptoms, potential disease cases, or resource shortages may produce data of unknown quality and veracity, which presents a significant limitation of

SMS data. Speaking about the anonymity of participants in a distributed SMS-based survey, one interviewee questioned, "*how robust is that data, how much faith and trust can you put in it?"* (Interview participant, mobile data platform service provider, 11 August 2015). Data produced by members of the public may not only be incorrect, it may also be biased which could negatively affect or put other members of the public at risk (Tracey *et al.*, 2015). One interview participant explained that during a phone-based campaign to identify Ebola cases in Sierra-Leone, many SMS messages were received from concerned people seeking information rather than from people with Ebola symptoms: "*the problem was how do you [quickly] sift out that kind of stuff from the real information that needs to be known right now"* (Interview participant, NGO volunteer, 28 April 2015).

In LMIC where mobile networks may already be weak, networks may get overloaded by a rapid increase in communications during a disease emergency (Muah *et al.*, 2014), which may limit the potential of SMS based data collection, however transmitting SMS is typically more stable and reliable than phone calls during an emergency (Revere *et al.*, 2014). Similarly, one interview participant was concerned that poor mobile connectivity particularly in rural areas could mean that resources are misdirected because "*Where you get random social media and SMS reporting of people expressing need, it tends to be the people that has the technology and can use it and shout the loudest the will come back asking for help. And they're not necessarily the ones in most need.*" (Inteview participant, NGO employee, 28 April 2015).

Similar to CDRs, a major challenge to using data received by SMS in disaster response is the time it takes to collate, clean and convert these data into formats useful for mapping, analysis, and subsequent decision-making. Data cleaning and processing is a time consuming task that needs to be completed prior to decision-making, and SMS based surveys do not always produce immediately actionable data, especially due to the anonymity of participants. As one participant noted, "*with these epidemics it has to be clean from the source, because you have no time to clean data*" (Interview participant, international aid organization, 7 May 2015). However, these SMS data processing challenges were refuted somewhat by another respondent: "*You get the [SMS] data, then you need to do some kind of analysis on it, and then really all you need is someone that understands something about statistics and analysing data….Which honestly should not be that difficult if you have monitoring and evaluation folks, or statistics folks, or economists, or researchers*" (Interview participant, mobile data platform service provider, 26 August 2015).

In general, raising awareness among the humanitarian sector about how, when and why to collect data using mobile phones would facilitate uptake. As one participant mentioned, "*There are few people in humanitarian or development organisations that I think actually understand the benefits that mobile data collection could bring to them and the limitations*" (Interview participant, mobile data platform provider, 26 August 2015). Yet widespread uptake of mobile data usage will require active support from donor, private sector (particularly mobile operator) and government (particularly mobile regulators) organisations to pay for training and project management, set up secure contracts to protect mobile users and mobile operators, provide incentives for mobile operators to provide access to mobile data or their networks in disaster contexts, and develop universal standards for data collection, processing and sharing.

3.5 Social and ethical challenges related to mobile data use

In addition to the issues outlined in the previous section, there are a number of associated ethical concerns that should be considered carefully by any organization wishing to use mobile data for disease disaster management. Perhaps most significantly, concerns about individual privacy were widely discussed in the literature and raised by interview participants, especially pertaining to the use of CDR data since these illustrate the locations, spatio-temporal movements, and communication patterns of individual phone users. Although the actual content of the communications are not included in CDR datasets, it is possible to identify sensitive details about individuals' activities by matching phone numbers from CDR datasets (if they are included) to publically available directories (Google, Yelp) and analysing patterns in the time of day and the locations from and to which calls were made (Mayer & Mutchler, 2014). When made available to researchers, CDRs are normally pseudonymized (names and numbers removed and replaced with a unique ID) and can be aggregated to the cell tower area to preserve privacy (Tatem *et al.*, 2014; Wesolowski *et al.*, 2014a), and encrypted using mathematical algorithms (GSMA, 2014). However, anonymization makes re-identification more difficult but still feasible by triangulation with online data that links individuals to geographic locations at specific times, such as geo-tagged social media (Cecaj *et al.*, 2016; de Montjoye *et al.*, 2014). Faced with these challenges in the Ebola response, GSMA released guidelines for mobile phone data use advising that CDRs remain secured and encrypted on the operator's server; all analysis completed within their premises; no attempts to de-anonymise IDs occur, and; only outputs (e.g. maps, models, aggregate statistics) should be shared with third parties (GSMA, 2014).

SMS data also raises significant privacy and confidentiality concerns. These systems are frequently designed to collect potentially sensitive information from participants, which – as in any survey or data collection exercise involving individuals – generally seek to retain participant anonymity. For SMS systems designed to collect information from people on resource needs or disease cases, location privacy will likely be compromised. For systems designed to collect symptom and vital sign data from the public, contributions will be highly sensitive personal information; as one participant described, the data collected via SMS-based systems for monitoring disease symptoms, although highly valuable for containing future outbreaks, is "*essentially an extension of a patient's medical records*" (interview participant, international organization employee, 29 April 2015). Yet, data transparency is crucial in these situations, and so knowing the persons and places in which contributions are coming from will be vital for decision-makers to establish its validity and credibility (Asokan & Asokan, 2015). For both SMS and CDR data, aggregation and anonymization are necessary privacy protections, however, they can seriously curtail their potential value. One participant noted, "*it limits us in terms of not having any information on who the people are … [w]e know for instance that certain segments of the population are much more likely to be malaria parasite carriers than others, and there are certain segments of the population that are much more mobile, and having that information could be very valuable to disease control programs and to responses to outbreaks*" (Interview participant, academic researcher, 10 November 2015).

The possibility of malevolence also has to be taken seriously, if systems are designed to rely on anonymous contributions. Humanitarian emergencies, and certainly communicable disease outbreaks, can be accompanied by a climate of fear and mistrust. Anonymous SMS reporting systems within these contexts could not only result in poor quality data, but could also provide a platform for fuelling community disharmony; as one participant noted, "*around something like Ebola, there is huge amounts of fear, there is huge amounts of you know, [people] looking over their shoulder, people reporting neighbours for things*" (Interview participant, mobile data platform provider). Similarly, cultural practices related to mobile phone use should be carefully considered such as gender issues, for instance; as one participant identified; "*we've heard stories if you're randomly [texting] people in communities, husbands will get jealous and angry … a woman gets a text on her phone from someone she doesn't know, you could be causing chaos*" (Interview participant, mobile data platform service provider, 11 August 2015).

The potential lack of representativeness of mobile data outlined in the previous section is more than just a barrier to use; there are also ethical issues that arise due to the uneven distribution of phones geographically and in populations. While those without a phone will not be subject to the potential privacy incursions outlined above, their voices will not be heard and they may miss out on the ability to shape agendas and resource decisions, or benefit from shared information (Buckee *et al.*, 2013; Shacklett, 2015). This is especially the case for SMS campaigns that provide critical information to users, such as the locations of nearby treatment facilities (Trad *et al.*, 2015).

There is considerable hype surrounding mobile technologies and big data, which present a challenge for those in positions of power and a concern for any organization that wishes to use them. Especially, expectations must be managed when communicating findings to stakeholders and the public: "*when people see the products that come out of these things, that gee-whiz factor sort of overwhelms and you know people will start focusing in on that. Unless you bring it into the context of everything else that's going on in an emergency, it can kind of swamp the other messages that need to get out*" (Interview participant, NGO volunteer, 28 April 2015). Aside from managing expectations, data users need to be aware of negative feedbacks that can arise from the introduction of a new data collection technique, especially in settings that have severely limited resources for disaster response. In the context of health emergencies, Higgins (2014) cautions that investing resources in new technologies and data sources can sometimes strip resources away from basic epidemiological analysis and disaster response. These examples suggest that, before employing SMS or CDR based data transfer systems, it is important to assess potential trade-offs and be upfront about any expected benefits or losses that might accrue to individuals or disaster response efforts.

Care must be taken when using CDR data and reporting findings to ensure information is not misleading or inaccurate. Sophisticated algorithms and data mining techniques are required to make sense of big data, and even then, conclusions are often made on the presence of correlations, of which the causal certainty is not well known. Although touted as the 'end of theory' – a new approach to science in which data speak for themselves (Anderson, 2008) – big data analytics should be informed by theoretical and specific domain knowledge and should involve human not just algorithmic and data-driven decision making (Miller & Goodchild, 2015), in order to make responsible and context relevant decisions: "*there is a lot of thinking that these data just on their own can solve a lot of problems but the importance of linking it with other data sources is incredibly valuable to provide that context and understand the data better*" (Interview participant, academic researcher, 10 November 2015). The effects of poor data analytics can be devastating in an emergency context. Vayena et al. (2015) describe how false identification of outbreaks and spread trajectories could result in inappropriate resource allocation decisions or harmful travel bans. However, the novelty of CDRs and big data analytics more generally means that the development of tools, software, and algorithms are in their infancy, and researchers need to develop and compare the accuracy, sensitivity and robustness of different CDR data analysis techniques to minimise the risk that these data can be misinterpreted.

In addition to individual level ethical concerns, the use of mobile phone data in disease disaster contexts could have serious societal consequences. The use of CDRs for surveillance and control as part of state security apparatuses cannot be ignored here, and their use in disaster management could be seen as an example of 'function creep', a process of normalization through data repurposing or recontextualization (Custers & Uršič, 2016) that serves to accelerate and strengthen the global corporate-government surveillance assemblage (Haggerty & Ericson, 2000; Murakami Wood, 2013; Taylor, 2015; Taylor & Broeders, 2015). Similarly, the surveillance potential of SMS needs to be better understood, 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). In disaster situations, using citizen-generated SMS data to define population groups based on risk categories is clearly problematic; when this process initiates different levels of assistance or intervention there is the possibility of engendering significant social harm or conflict. Arguably different rules apply in disaster situations, however the value of CDRs for monitoring movements and

SMS for gathering information from citizens must be carefully considered against threats to freedom and societal control (see Buscher *et al.*, 2013; Taylor, 2015). As Lyon (2007, p. 3) has argued, surveillance is a neutral concept (neither inherently good or bad); each specific implementation exists "somewhere on a continuum between *care* and *control*" (italics in original), and so it is vital that the potential for societal harm is carefully considered despite the ostensibly good intentions of disaster management and digital humanitarianism activities.

We use the results presented above to develop a decision-tree for mobile data use in disease-disaster management (Figure 3), which could help government, humanitarian and other organisations seeking information to inform disease disaster management decisions decide whether mobile phone data is an effective method to obtain the information they require in a specific institutional, socio-economic and geographic context.

Figure 3: Decision-tree for mobile phone data use in disease disaster management

4 Discussion and Conclusion

> "Huge potential, not yet fully refined, I think is the conclusion" (Interview participant, mobile data platform service provider, 11 August 2015)

Against the backdrop of potentially globalized health risks, including recurring communicable disease outbreaks such as Ebola and emerging threats such as Zika and Chikungunya, actively and passively produced mobile phone data could help information managers to fill information gaps, enabling more effective decision-making in all phases of disaster management, from prevention to recovery. This study attempts to fill knowledge gaps and advance our understanding regarding mobile phone data use in this context, and develops a typology and decision making tool to enable researchers and organizations involved in disease disaster management to ascertain how and if mobile data could be useful in future disaster decision-making. Evidence emerging from this research is likely to be relevant to the wider disaster context in many settings, not only to communicable disease emergencies.

In general, this study illustrates the rather different potential of the two mobile data types considered. While SMS data are limited in terms of the depth of data that can be obtained due to the constraints of the format and user-accepted survey length, SMS approaches can be very effective at rapidly extracting diverse types of information that is vital for disaster management, from resource needs to disease cases and symptoms. Combining the data collection functionality of SMS with the ability to broadcast vital information to dispersed citizens makes SMS a very powerful platform that could be effectively utilized by organizations and researchers of diverse technical ability, providing its limitations are properly understood and minimized. Organizations that are able to team up with MNOs to access their network and customer phone numbers – either directly or via an intermediary – may be better able to rapidly set up an SMS data collection/communication system. New tools such as RapidPro SMS are specifically designed to allow simple and quick development of SMS-based surveys at a range of scales.

Our results suggest that CDRs are limited in terms of the types of information that they can generate, yet their potential value for understanding population characteristics and dynamics in time-sensitive situations is unparalleled. Analysis of CDRs in disaster contexts can add huge value to existing data on population characteristics and dynamics. This is especially true for LMIC, which typically have inadequate or non-existent data infrastructures, yet which are frequently at the epicentre of communicable disease outbreaks. The current use of CDRs is, however, largely restricted to researchers and organizations that have big data analysis expertise, and the ability to engage in complex data access negotiations with regulators and MNOs. The future of CDR use in disease outbreaks and other humanitarian emergencies will therefore depend on the outcome of current discussions and negotiations aimed at accelerating access to these data, and on the transfer and uptake of technical knowledge within and across the humanitarian community. This includes establishing guidelines and protocols to ensure access and use of CDRs is considered acceptable by both MNOs and their customers.

Aggregated CDRs are unlikely to have truly *disruptive* potential in these contexts however (see Kitchin, 2014a). As McDonald (2016) describes, an unprecedented use of CDRs was undertaken in June 2015 in South Korea in the midst of a Middle East Respiratory Syndrome (MERS) outbreak; the government extracted personal information from mobile phone databases to pre-emptively quarantine and restrict the movements of 17,000 people based on algorithmically calculated infection probabilities. *Disaggregated*, *de-anonymized* CDRs could be transformative for such purposes as quarantine and contact tracing – vital activities that are often undertaken during contact-based disease outbreaks such as Ebola to contain spread – however more attention to the trade-off between social harm and the public good is needed around the use of CDRs in this format. This represents a crucial question for further academic and policy debate, despite some observers claiming that "data privacy and security … are largely irrelevant in the face of an epidemic outbreak" (Koch, 2016, p. 5).

Irrespective of the validity of this claim, the accumulation and use of big data has further societal costs beyond issues of privacy that must be more fully understood; highly relevant here are the ways in which power is being oriented away from citizens and governments towards corporations (boyd & Crawford, 2012; Thatcher *et al.*, 2016). The discussion of mobile data stakeholders here focuses on governments, MNOs, researchers, and disaster response organizations, yet omits an important group, the phone users themselves who actively or passively produce these data. Despite the data being explicitly about or produced by them, phone users are a little recognized stakeholder group in mobile data use debates, paralleling broader processes of disempowerment of 'data subjects' who neither control nor are well-informed about the uses of their data (see Kennedy & Moss, 2015; Leszczynski, 2015; Taylor, 2015; Zwitter, 2014). Shifting power relations are negatively impacting governments as well. The use of mobile phone data in LMICs presents significant opportunities for societal advancement through international development, humanitarian work, and disaster management, however the role of MNOs as 'data gatekeepers' is highly concerning, in an era of big data in which "power accrues to those who hold the most data" (Taylor & Broeders, 2015, p. 229). As mobile and other types of big data come to replace or stand in for datasets produced by public or non-governmental bodies (e.g. census or migration data), the growing ability of these private corporations to control how and what they are used for is a concern with significant ethical and democratic implications. Although this power shift has been documented in other areas of big data, notably the growing influence of IBM and other providers of 'smart city' solutions in shaping urban governance (Kitchin, 2014b; Townsend, 2013), the significant realized and potential value of mobile phone data highlighted in this paper illustrates the need for critical attention to similar powers of governance afforded to mobile network operators, an issue that may be especially problematic in LMIC.

Similarly, mobile data collection and communication implementations during disaster situations are frequently technocratic, top down, and centralized, which can produce uneven and damaging power relations between disaster relief organizations and distant 'needy' victims (see Burns, 2015). For the 2013-15 Ebola outbreak in West Africa, for example, there have been significant critiques of the communication strategies employed and their assumptions of behaviour and embedded social structures. Chandler *et al.* (2015) argue that many direct communication campaigns in West African countries sought to improve biomedical understanding of risks by targeted populations (e.g. hygienic practices) but that many campaigns were ineffective because they assumed local practices were deficient and were contributing to disease transmission. While it is unclear if these claims are substantiated, such critiques highlight the potential limitations of standardized protocols in some settings, as well as over-reliance on a single, centralized medium of delivery in engaging populations in managing disasters of many kinds. Hence, we stress the need for cultural and contextual awareness and sensitivity, especially in the face of complex emergency situations where poorly constructed campaigns can at best, waste resources, and at worst, undermine the disaster response effort.

Based on empirical findings from interviews and a scoping review, this paper develops a typology of mobile data sources, types, and end-uses and a decision-making tool which researchers and organizations involved in disease disaster management can use to ascertain at what phases of disaster management mobile data could meet their information needs, and what practical and ethical issues need careful attention before mobile data collection methods are attempted. The evidence suggests mobile data sourcing has a real potential to support disaster management efforts and there is growing interest among the humanitarian community in exploring its uses. Yet a range of issues still need to be addressed to secure widespread acceptance, access to, and informed use of mobile phone data in disease outbreaks and disaster management more generally. We recommend that future research efforts focus on: i) facilitating development of global mobile data usage guidelines, regulations and standards to provide rapid data access to approved organizations and for prescribed purposes, such as humanitarian crises; ii) documenting, comparing and contrasting existing tools and platforms that can assist in collection, management and dissemination of mobile data, in terms of their capabilities (input and output formats, computational options, compatibility with other information management tools), and usability (telecommunications infrastructure, skill type and level, time, accessibility, availability of training materials) as a minimum; iii) testing the validity and effectiveness of mobile data use in a wide range of technological and cultural contexts, including through in-depth monitoring and evaluation of active use cases; iv) develop quick and effective methods to identify and overcome gaps, anomalies and biases in mobile datasets.

Our findings show that sourcing data through mobile phones could significantly strengthen existing methods of data collection and transfer in all phases of disease-related disaster management. SMS and CDR data from mobile phones provide a means to rapidly gather baseline and situational information in data-scarce environments, and can significantly increase efficiencies, consistency and ease of data capture and transfer for institutions involved in disaster management. This study also finds that these benefits must be weighed against prevailing uncertainties regarding SMS and CDR data validity and representativeness, user privacy, and broader concerns regarding the potential for decisions based on mobile phone data to deepen societal power imbalances. Further research to match mobile data use options (data source, data type, end-use) to specific social, cultural, technological and disaster contexts is needed to accelerate the integration of mobile data sources into humanitarian response effort while ensuring these data complement and do not undermine other valuable data collection approaches.

References

Ahas, R., Aasa, A., Yuan, Y., Raubal, M., Smoreda, Z., Liu, Y., Ziemlicki, C., Tiru, M., & Zook, M. (2015). Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *International Journal of Geographical Information Science, 29*(11), 2017-2039.

Ahas, R., Miller, H. J., & Witlox, F. (2014). From the Guest Editors: Mobility, Communication, and Urban Space. *Journal of Urban Technology, 21*(2), 1-7.

Alexander, D. E. (2002). *Principles of Emergency Planning and Management.* Oxford, UK: Oxford University Press

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine, 16*.

Anderson, M. (2014). Ebola: how text-messaging is being used to help combat the outbreak. *The Guardian*. http://www.theguardian.com/global-development/2014/oct/16/ebola-text-messaging-combat-outbreak.

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19-32.

Asiimwe, C., Gelvin, D., Lee, E., Amor, Y. B., Quinto, E., Katureebe, C., Sundaram, L., Bell, D., & Berg, M. (2011). Use of an Innovative, Affordable, and Open-Source Short Message Service–Based Tool to Monitor Malaria in Remote Areas of Uganda. *The American Journal of Tropical Medicine and Hygiene, 85*(1), 26-33.

Asokan, G. V., & Asokan, V. (2015). Leveraging "big data" to enhance the effectiveness of "one health" in an era of health informatics. *Journal of Epidemiology and Global Health, in press*(0).

Aubrecht, C., Meier, P., & Taubenböck, H. (2015). Speeding up the clock in remote sensing: identifying the 'black spots' in exposure dynamics by capitalizing on the full spectrum of joint high spatial and temporal resolution. *Natural Hazards*, 1-6.

Bengtsson, L., Gaudart, J., Lu, X., Moore, S., Wetter, E., Sallah, K., Rebaudet, S., & Piarroux, R. (2015). Using Mobile Phone Data to Predict the Spatial Spread of Cholera. *Scientific Reports, 5*, 8923.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & von Schreeb, J. (2011). Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. *PLoS Med, 8*(8), e1001083.

Bharti, N., Lu, X., Bengtsson, L., Wetter, E., & Tatem, A. J. (2015). Remotely measuring populations during a crisis by overlaying two data sources. *International Health, 7*(2), 90-98.

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society, 15*(5), 662-679.

Buckee, C. O., Wesolowski, A., Eagle, N. N., Hansen, E., & Snow, R. W. (2013). Mobile phones and malaria: modeling human and parasite travel. *Travel Medicine and Infectious Disease, 11*(1), 15-22.

Burns, R. (2014). Moments of closure in the knowledge politics of digital humanitarianism. *Geoforum, 53*, 51-62.

Burns, R. (2015). Rethinking big data in digital humanitarianism: practices, epistemologies, and social relations. *GeoJournal, 80*(4), 477-490.

Buscher, M., Liegl, M., Rizza, C., & Watson, H. (2014). How to do IT more carefully: Ethical, Legal and Social Issues (ELSI) in IT supported crisis response and management. *International Journal of Information Systems for Crisis Response and Management, 6*(4).

Buscher, M., Wood, L., & Perng, S.-Y. (2013). *Privacy, Security, Liberty: Informing the Design of EMIS*. Paper presented at the Proceedings of the 10th International ISCRAM Conference, Baden-Baden, Germany.

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.* Washington, DC: Center for Global Development and The African Population and Health Research Center.

Chandler, C., Fairhead, J., Kelly, A., Leach, M., Martineau, F., Mokuwa, E., Parker, M., Richards, P., & Wilkinson, A. (2015). Ebola: limitations of correcting misinformation. *The Lancet, 385*(9975), 1275-1277.

Cinnamon, J., & Schuurman, N. (2013). Confronting the data-divide in a time of spatial turns and volunteered geographic information. *GeoJournal, 78*(4), 657-674.

Crawford, K., & Finn, M. (2015). The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal, 80*(4), 491-502.

Custers, B., & Uršič, H. (2016). Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection. *International Data Privacy Law, doi: 10.1093/idpl/ipv028.*

de Montjoye, Y.-A., Kendall, J., & Kerry, C. F. (2014). Enabling Humanitarian Use of Mobile Phone Data. *Issues in Technology Innovation, November*, 1-11.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., Blondel, V. D., & Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences, 111*(45), 15888-15893.

Douglass, R. W., Meyer, D. A., Ram, M., Rideout, D., & Song, D. (2015). High resolution population estimates from telecommunications data. *EPJ Data Science, 4*(1), 1-13.

Doyle, J., Hung, P., Farrell, R., & McLoone, S. (2014). Population Mobility Dynamics Estimated from Mobile Telephony Data. *Journal of Urban Technology, 21*(2), 109-132.

Fadiya, S. O., Saydam, S., & Zira, V. V. (2014). Advancing Big Data for Humanitarian Needs. *Procedia Engineering, 78*, 88-95.

FCS. (2012). *UK Standard for CDRs*. Beckenham, UK: Federation of Communication Services.

Gao, S., Liu, Y., Wang, Y., & Ma, X. (2013). Discovering spatial interaction communities from mobile phone data. *Transactions in GIS, 17*(3), 463-481.

Gilpin, L. (2014). How an algorithm detected the Ebola outbreak a week early, and what it could do next. *TechRepublic*. http://www.techrepublic.com/article/how-an-algorithm-detected-the-ebola-outbreak-a-week-early-and-what-it-could-do-next/.

Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: A research frontier. *International Journal of Digital Earth, 3*(3), 231-241.

Griffin, G. F. (2014). The Use of Unmanned Aerial Vehicles for Disaster Management. *GEOMATICA, 68*(4), 265-281.

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.
GSMA. (2015). *The Mobile Economy 2015*. London: Groupe Speciale Mobile Association.
Haggerty, K. D., & Ericson, R. V. (2000). The surveillant assemblage. *The British Journal of Sociology, 51*(4), 605-622.
Hern, A. (2014). Mobile phone records could help the fight against Ebola, study finds. *The Guardian*. http://www.theguardian.com/technology/2014/oct/29/mobile-phone-records-help-fight-against-ebola-study-finds.
Hey, A. J., Tansley, S., & Tolle, K. M. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft Research
Higgins, A. (2014). Fighting Ebola in Liberia with Technology. *Al Jazeera*. http://www.aljazeera.com/indepth/features/2014/11/fighting-ebola-liberia-with-technology-2014111584621373278.html.
Jacobsen, K. (2014). Big Data: How Cellphones Help Track Diseases. *Christian Science Monitor*. http://www.csmonitor.com/Technology/2013/0624/Big-data-How-cellphones-help-track-diseases.
Järv, O., Ahas, R., & Witlox, F. (2014). Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C: Emerging Technologies, 38*, 122-135.
Kennedy, H., & Moss, G. (2015). Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society, 2*(2).
Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography, 3*(3), 262-267.
Kitchin, R. (2014a). Big Data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1).
Kitchin, R. (2014b). The real-time city? Big data and smart urbanism. *GeoJournal, 79*(1), 1-14.
Koch, T. (2016). Fighting disease, like fighting fires: The lessons Ebola teaches. *The Canadian Geographer / Le Géographe canadien*, n/a-n/a.
Leszczynski, A. (2015). Spatial big data and anxieties of control. *Environment and Planning D: Society and Space, 33*(6), 965-984.
Lydersen, K. (2014). To Halt Ebola's Spread, Researchers Race for Data. *Discover Magazine*. http://blogs.discovermagazine.com/crux/2014/11/28/ebola-race-data/ - .VVxp01W6djB.
Lyon, D. (2007). *Surveillance Studies: An Overview*. Cambridge, UK: Polity.
Lyon, D. (Ed.). (2003). *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination*. London: Routledge.
Macharia, K. (2015). Africa: Is Big Data the Solution to Africa's Big Issues? *AllAfrica*. http://allafrica.com/stories/201502270779.html.
Madianou, M., Longboan, L., & Ong, J. C. (2015). Finding a Voice Through Humanitarian Technologies? Communication Technologies and Participation in Disaster Recovery. *International Journal of Communication, 9*, 19.
MapAction. (2016). Getting Help to Where its Needed Most. Retrieved 3 January 2016, from http://www.mapaction.org/about/about-us.html
Mayer, J., & Mutchler, P. (2014). MetaPhone: The Sensitivity of Telephone Metadata. Retrieved 28 Feb 2016, from http://webpolicy.org/2014/03/12/metaphone-the-sensitivity-of-telephone-metadata/
McDonald, S. M. (2016). *Ebola: A Big Data Disaster - Privacy, Property, and the Law of Disaster Experimentation*. Bengaluru: Centre for Internet and Society.
Meier, P. (2015). *Digital Humanitarians: How Big Data is Changing the Face of Humanitarian Response*: CRC Press.
Miller, H. J. (2010). The data avalance is here: Shouldn't we be digging? *Journal of Regional Science, 50*(1), 181-201.
Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal, 80*(4), 449-461.

Muah, S., Kochi, E., & Fabian, C. (2014). How The Tech Sector Can Help Stop Ebola. *TechCrunch*. http://techcrunch.com/2014/10/29/tech-ebola/.
Murakami Wood, D. (2013). What is global surveillance? Towards a relational political economy of the global surveillant assemblage. *Geoforum, 49*, 317-326.
Mushayi, W. (2015). Making schools safer in Sierra Leone during the Ebola outbreak. Retrieved 6 May 2015, from http://blogs.unicef.org/2015/05/06/making-schools-safer-in-sierra-leone-during-the-ebola-outbreak/
O'Donovan, J., & Bersin, A. (2015). Controlling Ebola through mHealth strategies. *The Lancet Global health, 3*(1), e22.
Oliver, N. (2013). Combating global epidemics with big mobile data. *The Guardian*. http://www.theguardian.com/media-network/media-network-blog/2013/sep/05/combating-epidemics-big-mobile-data.
Pew Research Center. (2015). *Cell Phones in Africa: Communication Lifeline*. Washington, DC: Pew Research Center.
Pu, C., & Kitsuregawa, M. (Eds.). (2013). *Technical Report No. GIT-CERCS-13-09 - Big Data and Disaster Management: A Report from the JST/NSF Joint Workshop* Georgia Institute of Technology
Revere, D., Schwartz, M. R., & Baseman, J. (2014). How 2 txt: an exploration of crafting public health messages in SMS. *BMC Research Notes, 7*, 514.
Roche, S., Propeck-Zimmermann, E., & Mericskay, B. (2013). GeoWeb and crisis management: Issues and perspectives of volunteered geographic information. *GeoJournal, 78*(1), 21-40.
Scepanovic, S., Mishkovski, I., Hui, P., & Ylä-Jääski, A. (2015). Mobile Phone Call Data as a Regional Socio-Economic Proxy Indicator. *PLoS ONE, 10*(4), e0124160.
Shacklett, M. (2015). Fighting Ebola with a holistic vision of big data. Retrieved 12 April 2015, from http://www.techrepublic.com/article/fighting-ebola-with-a-holistic-vision-of-big-data/
Shanley, L., Burns, R., Bastian, Z., & Robson, E. (2013). Tweeting up a storm: the promise and perils of crisis mapping. *Photogrammetric Engineering and Remote Sensing, 79*(10), 865-879.
Shelton, T., Poorthuis, A., Graham, M., & Zook, M. (2014). Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of 'big data'. *Geoforum, 52*, 167-179.
Tatem, A., Qiu, Y., Smith, D., Sabot, O., Ali, A., & Moonen, B. (2009). The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. *Malaria Journal, 8*(1), 287.
Tatem, A. J. (2014). Mapping population and pathogen movements. *International Health, 6*(1), 5-11.
Tatem, A. J., Huang, Z., Narib, C., Kumar, U., Kandula, D., Pindolia, D. K., Smith, D. L., Cohen, J. M., Graupe, B., Uusiku, P., & Lourenço, C. (2014). Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria Journal, 13*, 52-52.
Taylor, L. (2015). No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *Environment and Planning D: Society and Space*, 0263775815608851.
Taylor, L., & Broeders, D. (2015). In the name of Development: Power, profit and the datafication of the global South. *Geoforum, 64*, 229-237.
Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016). Data Colonialism Through Accumulation by Dispossession: New Metaphors for Daily Data. *Environment and Planning D: Society and Space, DOI: 10.1177/0263775816633195*.
The Economist. (2014). Waiting on Hold: Mobile-phone records would help combat the Ebola epidemic. But getting to look at them has proved hard. *The Economist*. http://www.economist.com/news/science-and-technology/21627557-mobile-phone-records-would-help-combat-ebola-epidemic-getting-look.
The Guardian. (2014). Ebola: how text-messaging is being used to help combat the outbreak. Retrieved 1 May 2015, from http://www.theguardian.com/global-development/2014/oct/16/ebola-text-messaging-combat-outbreak.Accessed
Thompson, S., Altay, N., Green III, W. G., & Lapetina, J. (2006). Improving disaster response efforts with decision support systems. *International Journal of Emergency Management, 3*(4), 250-263.
Townsend, A., M. (2013). *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. New York: WW Norton and Company.

Tracey, L. E., Regan, A. K., Armstrong, P. K., Dowse, G. K., & Effler, P. V. (2015). EbolaTracks: an automated SMS system for monitoring persons potentially exposed to Ebola virus disease. *Euro Surveillance, 20*(1).
Trad, M.-A., Jurdak, R., & Rana, R. (2015). Guiding Ebola patients to suitable health facilities: an SMS-based approach. *F1000Research, 4*, 43.
UN International Strategy for Disaster Reduction. (2009). *UNISDR Terminology on Disaster Risk Reduction*. Geneva: United Nations ISDR.
Vasilescu, L., Khan, A., & Khan, H. (2008). Disaster management cycle–A theoretical approach. *Management & Marketing*(1), 43-50.
Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical Challenges of Big Data in Public Health. *PLoS Comput Biol, 11*(2), e1003904.
Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., & Mehl, H. (2007). Satellite image analysis for disaster and crisis-management support. *Geoscience and Remote Sensing, IEEE Transactions on, 45*(6), 1520-1528.
Wesolowski, A., Buckee, C. O., Bengtsson, L., Wetter, E., Lu, X., & Tatem, A. J. (2014a). Commentary: Containing the Ebola Outbreak - the Potential and Challenge of Mobile Network Data. *PLoS Currents, 6*, ecurrents.outbreaks.0177e0177fcf52217b52218b634376e634372f634373efc634375e.
Wesolowski, A., Stresman, G., Eagle, N., Stevenson, J., Owaga, C., Marube, E., Bousema, T., Drakeley, C., Cox, J., & Buckee, C. O. (2014b). Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. *Sci. Rep., 4*.
Whitford, H. M., Donnan, P. T., Symon, A. G., Kellett, G., Monteith-Hodge, E., Rauchhaus, P., & Wyatt, J. C. (2012). Evaluating the reliability, validity, acceptability, and practicality of SMS text messaging as a tool to collect research data: results from the Feeding Your Baby project. *Journal of the American Medical Informatics Association, 19*(5), 744-749.
WorldPop. (2014). Ebola. Retrieved 12 April 2015, from http://www.worldpop.org.uk/ebola/
Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Lu, F., Chen, J., Fang, Z., & Li, Q. (2016). Another Tale of Two Cities: Understanding Human Activity Space Using Actively Tracked Cellphone Location Data. *Annals of the American Association of Geographers, 106*(2), 489-502.
Yang, C., Yang, J., Luo, 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*(8), 619-623.
Young, S. D. (2015). A "big data" approach to HIV epidemiology and prevention. *Preventive Medicine, 70*(0), 17-18.
Zhao, Z., Shaw, S.-L., Xu, Y., Lu, F., Chen, J., & Yin, L. (2016). Understanding the bias of call detail records in human mobility research. *International Journal of Geographical Information Science*, 1-25.
Zhong, S., & Bian, L. (2016). A Location-Centric Network Approach to Analyzing Epidemic Dynamics. *Annals of the American Association of Geographers, 106*(2), 480-488.
Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy, 2*(2), 7-33.
Zoomers, A., Gekker, A., & Schäfer, M. T. (2016). Between two hypes: Will "big data" help unravel blind spots in understanding the "global land rush?". *Geoforum, 69*, 147-159.
Zwitter, A. (2014). Big Data ethics. *Big Data & Society, 1*(2), 1-6.