A method for near real-time surveillance of hepatitis A and E cases in Ahmedabad, India

By

Carl Hughes

Department of Geography
McGill University, Montreal
April 2016

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Arts

© Carl Hughes 2016

Table of Contents

List of Tables

List of Figures

Abstract

Disease surveillance is the continuous monitoring of the distribution and trends of disease cases through systematic collection, consolidation, and evaluation of case reports (Langmuir, 1963). Geo-spatial disease surveillance, a sub-focus of disease surveillance, relies on the principle of tracking the location of cases through space and time to identify clusters of disease. Contemporary methods for disease surveillance increasingly use information communication technologies (ICTs) to increase the accuracy, precision, quality, and utility of disease monitoring (Cinnamon & Schuurman, 2010).

Though much attention is paid to improving public health services in countries like India, substantial challenges remain, including lack of personnel or structured reporting mechanisms for the surveillance of waterborne diseases and outbreak detection. Presently, the city of Ahmedabad uses a "manual follow-up" based reporting system under the framework of the Integrated Disease Surveillance Programme (IDSP) to monitor relevant diseases, including hepatitis A and E. In this thesis, I evaluate the existing geo-spatial disease surveillance capability in Ahmedabad, and assess a possible alternative for improved outbreak detection using a case-study of hepatitis A and E cases.

First, I present a comprehensive literature review on urban health inequities in Indian cities, and the mechanisms through which they manifest, to justify research in disease surveillance methods. Next, I use guidelines proposed by the US Centre for Disease Control (CDC) to create a system evaluation to assess the IDSP in terms of its capabilities in geo-spatial monitoring of waterborne disease for outbreak detection. The results of this evaluation suggest the IDSP cannot quickly or accurately identify the locations of outbreaks in a near-real time. Finally, based on this evaluation, I develop, prototype, and evaluate – using hepatitis A and E case data − a new method for near real-time disease surveillance with respect to its ability to address some of the main challenges facing the IDSP. The evaluation of this prototype suggests a mobile based data collection tool could be used to help improve outbreak detection in terms of accuracy, precision, and timeliness. I identify costs, scalability, training, and technical capacity as barriers to widespread adoption of such a tool.

Résumé

La surveillance des maladies est la surveillance continue de la répartition et les tendances des cas de maladie à la collecte systématique, la consolidation et l'évaluation des rapports de cas (Langmuir, 1963). La surveillance des maladies géo-spatiale, un sous-type de surveillance de la maladie, repose sur le principe de suivre de l'emplacement des cas à travers l'espace et le temps pour identifier les groupes de maladies. Les méthodes modernes pour la surveillance des maladies utilisent de plus en plus les technologies de communication de l'information (TIC) pour augmenter la précision, la qualité et l'utilité de la surveillance de la maladie (Cinnamon & Schuurman, 2010).

Bien que beaucoup d'attention ait été accordée à l'amélioration des services de santé publique dans des pays comme l'Inde, de nombreux défis subsistent, notamment concernant le manque de personnel ou de systèmes de surveillance structurés pour la détection de maladies d'origine hydrique et des épidémies. Actuellement, la ville d'Ahmedabad utilise un système manuel basé sur le suivi des cas ayant comme fondation l'Integrated Disease Surveillance Programme (IDSP) pour surveiller les maladies comme l'hépatite A et E. Dans cette thèse, j'évalue la capacité de surveillance des maladies géospatiales existante à Ahmedabad et propose une solution possible pour une meilleure détection des épidémies.

D'abord, j'examine la littérature sur la santé en milieu urbain, les injustices et les mécanismes par lesquels ils se manifestent pour justifier la recherche sur les méthodes de surveillance des maladies. Ensuite, avec l'aide des recommandations d'évaluation de systèmes développées par le Centre for Disease Control (CDC), l'IDSP sera évalué en termes de son efficacité à surveiller et détecter l'évolution de la propagation de maladies d'origine hydrique. Les résultats de cette évaluation suggèrent que le IDSP ne peut pas rapidement détecter l'évolution de la propagation de maladies d'origine hydrique avec précision dans un temps quasi-réel. Enfin, en lien avec cette évaluation, une nouvelle méthode de surveillance de la maladie en temps quasi réel sera décrite et mise à l'essai pour adresser certains des principaux défis confrontés par l'IDSP. Ce prototype a été testé avec succès sur le terrain afin de simuler un déploiement opérationnel. L'évaluation de ce prototype propose un outil de collecte de données mobiles pourrait être

utilisé pour aider à améliorer la détection des épidémies en termes de précision, et la rapidité. J'identifie les coûts, l'évolutivité, la formation et les capacités techniques comme des obstacles à l'adoption généralisée.

Acknowledgements

This thesis would not have been possible without the generous personal, professional, and academic support I have received throughout the past two years from my friends, family and colleagues. I am extremely thankful to my supervisor, Dr. Raja Sengupta, for pushing me to expand my boundaries academically and encouraging me to pursue my interests. Thank you for your guidance. I would like to thank my committee member, Dr. Lea Berrang Ford for your extensive insight into the field of health Geography and invaluable direction you provided for the thesis writing process.

I appreciate the mentorship and generosity provided to me from Drs. Vinayak Naik, Deepak Saxena, and Kranti Vora. I am grateful for the tremendous assistance you have provided me in shaping my research and hosting me during my time in India. Thank you Dr. S. Yasobant, Ashish Upadhyay, Zuned Mansuri, and Pankaj Yadav for welcoming me to Indian Institute of Public Health Gandhinagar and your tireless efforts in assisting me with my fieldwork. Thank you Ahmedabad Municipal Corporation (AMC) for allowing my research to proceed and providing me with your datasets.

Thank you to my lab-mates: Dipto Sarkar, Suthee Sangiambut, Matthew Tenney, Jin Xing, Drew Bush, Genèvieve Reid, Alex Taciuk, and Ana Brandusescu, and Dr. Renee Sieber, and Jing Hoon Teo for providing such a stimulating research environment. I appreciate the intellectual discussions, critical-feedback, proof-reading, collaboration, and general support. Thank you Dr. Gwyn Campbell, Lori Callaghan, Dr. Erik von Dongen and the entire Indian Ocean World Centre and Social Sciences and Humanities Research Council (SSHRC) for your generous financial and logistical support. Thank you Mr. Warren, Mitacs and the Geography Department at McGill University (GREAT Award and Grad Excellence Award) for the generous funding in support of my research. Thank you to my family, Brett, Susan, and John for your continuous encouragement and support. You always have my best interests in mind and I appreciate all the useful advice you have shared with me in pursuit of my goals.

Preface and Contribution of Authors

This thesis, composed of two core chapters (chapter 3 and 4), is written in manuscript format (i.e. both chapters will be submitted to peer-reviewed journals). For both articles, I was responsible for developing the research objectives, formulating the analytical framework and methodology, carrying out the analysis, interpreting the results, and writing the manuscripts. My M.A. committee, Dr. Raja Sengupta and Dr. Lea Berrang Ford, assisted with the initial development of my research questions. In addition, my collaborators from India, Dr. Deepak Saxena, helped in the organization of field workers for the prototype field testing, and helped design the case report form; and Dr. Vinayak Naik helped with the selection of the technical components for the prototype (data collection and database).

Manuscript I (Chapter 3):

"Evaluation of the Integrated Disease Surveillance Programme (IDSP) in Ahmedabad for outbreak detection" by Carl Hughes, Raja Sengupta, Deepak Saxena, and Lea Berrang-Ford.

Manuscript II (Chapter 4)

"Examining the role of mHealth for spatial disease surveillance" by Carl Hughes, Raja Sengupta, Deepak Saxena, and Vinayak S. Naik.

Chapter 1. Introduction

1.1 Research context

Urban health outcome disparities in Indian cities have been attributed to the failure of the Indian state to provide equitable water, sanitation and hygiene (WASH) services (Chaplin, 2011). However, this assertion is a highly contentious and politically charged topic (Saravanan et al., 2015). Contagious, and often fatal, diseases like cholera, viral hepatitis, typhoid and acute gastroenteritis (AGE) are endemic to urban cities in India, including Ahmedabad, with outbreaks occurring frequently in slum areas (Chauhan et al., 2010; Saravanan, 2013; Vyas et al., 2010). From a public health perspective, this means resources spent on crisis response often exceed expenditure for preventive action (Chaplin, 2011). In the long-term, preventive measures such as better and more equitable water infrastructure, and increased health education may be most beneficial. In the short-term, improving outbreak detection tools can help improve response activities (Carneiro & Mylonakis, 2009; Jajosky & Groseclose, 2004; Kant & Krishnan, 2010). Therefore, it is important that public health response interventions be implemented in parallel with surveillance resources to ensure the best response strategies.

Officially, drinking water in the Indian city of Ahmedabad is provided through public standpipes under the Basic Urban Service Program and through water delivery trucks (Herma, 2013). Up to 25 percent of Ahmedabad residents use illegal connections from public water pipes for drinking and waste water (Herma, 2013). The widespread illegal connections and unauthorised repair work in Ahmedabad is reflective of other major urban cities in India and throughout developing countries (Rakodi, 2000; Truelove, 2011). Many upscale villas and apartment complexes also use illegal connections that contravene city regulations (Chaplin, 2011). The main difference is that for the informal settlements, low quality pipes are used and laid by untrained labourers. It is also very rare for appropriate water filtration devices to be used in slum areas (Saravanan, 2013).

This results in frequent and widespread contaminations of drinking water, which in turn is a predominant source for waterborne viral hepatitis transmission (Chaplin, 2011). The problem is further compounded by limited infrastructure provided by the city where the few repairs that are carried out are delayed and of poor quality. Official work sites are often abandoned and left uncovered (Figure 1), which can deteriorate quickly, contributing to the mixing of drinking and sewage flows. The undocumented nature of this work means city officials cannot keep track of the locations of these sites. It is difficult to identify high-risk areas for disease because of the link between viral hepatitis and illegal infrastructure work which makes surveillance of outbreaks challenging.

Figure 1. Abandoned and uncovered water pipe repairs in Ahmedabad

A component of public health, disease surveillance entails the continuous monitoring of the distribution and trends of disease cases through systematic collection, consolidation, and evaluation of case reports (Langmuir, 1963). Additionally, it involves making these data, and consequent interpretations, available to relevant stakeholders concerned with the public health of a community in a timely manner. Relevant information access for public health decision makers is difficult in resource-scarce settings, such as India (Suresh, 2008). The use of mobile phones and other information communication technologies (ICTs) for data transmission has been suggested to reduce this burden by increasing connectivity to stakeholders (Yadav, 2010). Documentation on the practical use of such technology for disease surveillance remains relatively experimental in the literature.

Surveillance seeks to achieve two primary objectives: first to trigger disease-control interventions, and second to substantiate long-term trends in disease occurrence (Hashimoto et al., 2000; Thacker et al., 1983). The intention is that meeting these objectives will minimize exposure to risk for individuals, reducing morbidity and mortality in the population. In practice, disease surveillance is often achieved through government-mandated reporting mechanisms, whereby physicians, laboratories, or hospital authorities are required to report cases to a centralized surveillance authority. This method of disease surveillance is known as passive surveillance (Berkelman et al., 1994). Conversely, active surveillance is achieved through public health agencies directly contacting physicians, laboratories and hospitals, as well as schools or other organizations to “search” for new cases, often during an existing outbreak or periods with high incidence rates. Approaches that use ICTs are being tested as a core component of surveillance for this purpose owing to the potential for simple, reliable, and rapid data collection (Cinnamon & Schuurman, 2010; Kant & Krishnan, 2010). Once collected, case reports can analytically infer the presence of an outbreak through statistical methods. Unkel et al. (2012) identify statistical methods incorporating spatial information as one such approach. Specific methods are presented and discussed in Appendix A.

Geo-spatial disease surveillance, a sub-focus of disease surveillance that uses geographic analysis methods, is an important aspect of public health, especially in the context of waterborne disease (Unkel et al., 2012). The probability of transmission of waterborne disease decreases with distance from the source of infection. Therefore, waterborne diseases are influenced by the spatial position and dispersion of pathogens located in fresh water (Ostfeld et al., 2005). Outbreak detection methods that use spatial information can potentially detect localised outbreaks, or identify spatial patterns of disease that point to sources of exposure or risk. Such spatial methods typically require some indicator of the distance between cases, or their location. This information is difficult to attain without the use of reliable and accurate spatial data which can be facilitated through the use of GPS devices or highly accurate paper maps. The relative cost and direct integration of GPS technology in mobile devices (Krishna et al., 2009) make collecting these kinds of data through ICTs a potentially attractive option.

Knowing the location of individual cases can be used to identify spatial hot spots or clusters of cases which can help trace a contamination source (Zhou et al., 2013). Clusters of cases can be compared against a measure of spatial randomness within a surveillance unit to determine if the rate of disease at the cluster location is an anomaly. Areas with significantly higher incidence than the rest of the study area can be categorized as such through the use of hot spot analysis (Getis & Ord, 1992; Ord & Getis, 1995). In disease surveillance, hot spots are one spatial method used to determine if and where an isolated outbreak of disease exists within a pre-defined area (Unkel et al., 2012). Hot spot analysis requires that cases contain some geographic reference, for example city ward or street. They also assume georeferenced base data/maps against which the references can be geocoded (e.g., ward divisions or a street network). Thus, spatial statistical analysis can be used to help identify the location of an outbreak of disease, and potentially be used to help locate a spatial risk factor associated with the outbreak. This information can help guide administrators in the design and implementation of a targeted response to reduce risk exposure. Such a method, however, is technology intensive, and challenging to achieve without a geographic information system (GIS) or spatial analysis software package (Chen et al., 2011; Elliott & Wartenberg, 2004).

Geo-spatial disease surveillance tools have been used by public health organizations to better understand the spatial properties of disease and health outcomes, helping to inform public health strategies, both long and short-term. Elias et al. (2006) found that through spatiotemporal analysis of meningococcal cases in Germany, strain and region specific outbreaks were better assessed for public health prevention and containment efforts. Similarly, a spatial analysis of cholera cases during the 2010 epidemic in Haiti was able to help locate the contamination source for several significant clusters across the country (Piarroux et al., 2011). Gao et al. (2008) report on the benefits from using an online and interactive geographic information system (GIS) for infectious disease surveillance monitoring in New Brunswick and Maine. They found that the use of a spatial platform improved efficiency and effectiveness of disease surveillance through strengthening the capacity to do analysis, visualization of cases, and cross-collaboration with different stakeholders.

Further, timely reporting of new cases with accurate data can help strengthen a public health response (Zhou et al., 2013). To identify an outbreak in its early stages, cases are ideally reported to the central authority in place (Langmuir, 1963), or as close as possible, with the actual identification of new cases (from a diagnosing authority). Methods exist to monitor proxy indicators for disease to predict an emerging outbreak at the very early stage through the use of syndromic case reports (Mandl et al., 2004) or predictive analytics (Siegel, 2013; Vergu et al., 2006), for example, tracking the sale of certain medicines, or monitoring web searches for specific symptoms. These technologies are relatively new and their utility in developing country settings is still being tested (Mykhalovskiy & Weir, 2006).

Systemic delays in reporting of new cases are often related to the continued spread of an outbreak due to ongoing exposure to risk factors for otherwise healthy persons. Zhou et al. (2013) argue that reducing reporting delays from diagnosis to the reporting authority has a direct positive impact on the ability to control outbreaks. A measure of effectiveness for a surveillance system lies in the reduction of delays. Overall, this will improve the decision-making process and cost-efficiency of public health by lowering the cost of disease surveillance and re-allocating resources towards infrastructure issues or response efforts. In countries like the Netherlands and United States, waterborne disease diagnoses are required to be reported within one day (California, 2011; Reijn et al., 2011). This quick turnaround is realized due to readily available digital reporting systems enabling electronic case reporting (Krause et al., 2007). With electronic medical records (EMR), email, and mobile based health tools, submission of cases to a centralized authority is almost instantaneous and can allow for data verification.

This thesis is largely a response to calls in the literature to evaluate the opportunities and challenges for mHealth in addressing public health issues, in particular, spatial disease surveillance (DeSouza et al., 2014; P. N. Mechael, 2009). mHealth, the exploitation of mobile technology for health applications (Istepanian et al., 2004; Istepanian & Lacal, 2003; Istepanian et al., 2006), is one ICT approach that has been recommended for improving disease surveillance methods. The use of mHealth systems has been put forth because of the rapid and widespread diffusion of mobile phones coupled with low hardware and operating costs. In theory, mobile phones provide flexibility, communicative functions, and connectivity that can be used for disease prevention, surveillance and management, (Déglise et al., 2012) through rapid, reliable, and cost effective data collection, storage, and transmission. Although many mHealth projects are currently being used in India, mHealth as a public health and disease surveillance tool is currently largely experimental and the spatial functionality of mobile phones remain unleveraged in existing mHealth systems. Furthermore the use of ICTs for public health in developing countries is challenged by the need for technical proficiency and cost (Cinnamon & Schuurman, 2010; Kant & Krishnan, 2010).

This research considers the argument made by Butler (2006) that due to the continued burden of communicable disease, developing countries require a fundamental overhaul of disease surveillance to capitalize on real-time data updates as well as to become more cost-effective (Cinnamon & Schuurman, 2010); Mehl et al. (2014) propose the use of spatial technology in mHealth applications as one possible approach. The importance of such technologies may increase the capacity in developing countries, such as India, for carrying out spatial disease analysis. Suresh (2008) asserts that due to the high burden of disease in India, it has been difficult to detect, diagnose and control outbreaks of disease until the magnitude is quite large. By introducing better and automated surveillance mechanisms, interventions to target risk factors can be implemented swiftly and accurately with regards to temporal and spatial location. It is possible, however, that the technological dependency and lack of validation for such methodologies will pose operational challenges in a real deployment, potentially further complicating disease surveillance in a resource-scarce setting. This thesis aims to contribute evidence towards these discussions by critically reviewing the existing disease surveillance system in Ahmedabad and evaluating the potential for mHealth to address existing weaknesses through an operational pilot of one mHealth system.

1.2 Research questions and objectives

Question 1. What are the strengths and challenges of the existing disease surveillance system in Ahmedabad, India, for the control of waterborne disease?

In Ahmedabad, the impact of waterborne disease outbreaks are wide-reaching in terms of number of cases, as well as the economic and livelihood impacts of patients and their communities from substantial healthcare costs and lost employment wages. Ahmedabad experiences frequent outbreaks of waterborne disease, often attributed to the mixing of drinking water and sewage flows from illegal tampering of the city's water infrastructure. Due to the challenge in monitoring this type of risk factor, the location of outbreaks can be unpredictable. I hypothesize that surveillance of waterborne disease cases can be used to direct public health resources where and when appropriate. Poor surveillance methods may inhibit the ability to locate an outbreak and expose otherwise healthy persons to the contamination source. This question seeks to evaluate the ability for the existing surveillance program in Ahmedabad, the IDSP, to respond to outbreaks of waterborne disease. Chapter 3 will answer this question.

Question 2. Can spatial mHealth tools be leveraged for overcoming operational issues in waterborne disease control of low-resource areas? A case-study using viral hepatitis data.

The literature on mHealth suggests it has the potential to initiate disease-control responses through automated hot spot detection to improve disease surveillance. A prototype for near real-time geo-surveillance of waterborne disease, using viral hepatitis cases, is presented. Viral hepatitis was selected due to the high frequency and cyclical nature of local outbreaks attributed to contaminated water. This prototype is designed to address some of the existing challenges facing the IDSP in Ahmedabad and provide insight into the potential for mHealth to address some of the existing challenges in disease surveillance. This case study was conducted during an outbreak of viral hepatitis and the results from field-testing will be used to examine the applicability and challenges of spatial and open source mobile phone-based tools to improve disease surveillance in low-resource settings. Chapter 4 answers this question.

1.3 Methodology

The methodology of this thesis is twofold. To answer Question 1, I evaluate the local implementation of the Integrated Disease Surveillance Programme (IDSP) in Ahmedabad. The framework I use was selected to provide an objective assessment of the current surveillance system. The evaluation uses four months' worth of case reports collected during regular surveillance by community health workers, and provided to me by the Ahmedabad Municipal Corporation (AMC). To answer Question 2, I use a mobile phone-based data collection tool, web-based geographic information systems (GIS), open source digital database, and spatial analysis algorithms to design an automated analysis system for disease surveillance using mHealth components. This technology was tested during an outbreak of viral hepatitis in Ahmedabad to simulate an operational deployment. The results from field-testing of the prototype are then used to answer Objective 2.

1.4 Thesis structure

This thesis is divided into five chapters. In Chapter 2, I review the relevant peer-reviewed and grey literature on public health and the impact of waterborne disease in urban India. I then review existing mHealth technologies and their limitations as a public health tool, with a focus on developing countries. This review contextualizes health services and outcomes for the urban poor, and identifies efforts and challenges in addressing public health priorities. This chapter builds the case for this thesis' research questions to evaluate disease surveillance methods and investigate the potential for mHealth in monitoring waterborne disease.

In Chapter 3 I present an evaluation of the IDSP in Ahmedabad. First, I describe the purpose and structure of the IDSP and make the case for system evaluation. I then evaluate core system attributes from a review of grey literature, and examining a dataset of 2,194 viral hepatitis case reports collected by community health workers and link workers under the IDSP in Ahmedabad. These data provide evidence to assess the performance of the IDSP. The core system attributes that are evaluated include: simplicity, flexibility, data quality, acceptability, sensitivity, positive predictive value, representativeness, timeliness, and stability. Although this evaluation focuses on the local deployment of the IDSP, the national framework is briefly considered throughout the thesis to provide context. Based on this evaluation, I make conclusions about the overall ability for the IDSP to carry out geo-spatial disease surveillance and outbreak detection. I found that data quality and timeliness are two key system attributes that inhibit the overall ability to fulfill the local public health goals.

In Chapter 4, I present the design and results of a case study and evaluation of a prototype surveillance system. First, I describe the key objectives the prototype seeks to achieve and their relevance to robust disease surveillance systems. The specific objectives for the prototype are 1) ensuring rapid reporting of case reports for analysis, 2) ensuring data security and confidentiality in a digital database, 3) collecting accurate and standardized case reports, which include GPS coordinates, 4) using free and open source technology where available, 5) and automating spatial analysis, and subsequent visualization of results, of the distribution of disease to identify hot spots. I then present a detailed methodology that was used to create the system. The case study is then evaluated from field-testing based on its ability to meet the given objectives. I discuss how the prototype compares with the existing system. I compare relevant system attributes appropriate for prototype system evaluation, mainly data quality and timeliness, with the IDSP. Results from this case study suggest data quality, timeliness, and cost-effectiveness can be improved using open source mHealth tools. Challenges that emerged include: user training, technical dependency, limited analytical power, and scalability.

Finally, Chapter 5 concludes the thesis and draws together the most important findings from Chapter 3 and 4. I discuss how this thesis provides insights into the research questions and makes a contribution to the literature on disease surveillance and mHealth.

Chapter 2. Literature Review and Discussion

2.1 Introduction

In this chapter, I bring together relevant literature from public health, spatial methods for disease surveillance, and mHealth. I focus the review on highlighting calls for, and challenges with, novel approaches to public health due to the nature of waterborne disease which is related to space and proximity to pathogens, and existing health inequities in urban areas of Ahmedabad. First, I introduce and review the concept of health in the context of international development and attempt to describe current health challenges facing the city of Ahmedabad. Secondly, I review the mechanisms in which some of these issues have emerged. Next, I describe the field of mHealth and comment on its contributions and critiques *vis-à-vis* public health and geo-spatial disease surveillance in developing areas. Finally, I conclude the literature review by drawing out the contribution this thesis will make by answering the given research questions.

2.2 Literature Review

2.2.1 Global Health

In what is known as the Alma-Ata declaration (1978), the International Conference on Primary Health Care called for the international community to protect and promote health for all in the developing world (World Health Organization, 1978). Since then, the international development community has widely accepted that good health for everyone must be a priority for national governments, intergovernmental organizations, and civil society (Collier & Dollar, 2001; Jamison et al., 2006; Jamison & Mosley, 1991; Sachs & McArthur, 2005). What constitutes health and good health remains, however, unclear (Patrick et al., 1973; Saracci, 1997) and has been defined in terms of longevity, quality of life, and the mere absence of disability and pain (Larson, 1999).

The World Health Organization (WHO) (1978) broadly defines health as being the “state of complete physical, mental and social well-being, and not merely the absence of disease or infirmity”. This working definition fits well within a capabilities approach to development as it captures both the bio-medical aspects of health, or freedom from disease, in addition to a social component of general well-being. Sen (1999) reaffirms the flexible and broad concept of health in his definition of development as freedom, specifically, as the “process of expanding real freedoms that people enjoy”. Fitting into the capability approach framework, good health is seen as a resource, capability, and function involved in living a good life with reason to value.

2.2.2 Disease burden on urban poor in Ahmedabad

Since independence, India has faced a significant gap in health outcomes for the urban poor (Agarwal, 2011). The uneven burden of disease on the poor is largely attributed to the high prevalence of neglected disease, or “diseases of poverty” (e.g. dengue, Chagas disease, leprosy) (Manderson et al., 2009), inadequate levels of public health expenditure, and limited health facilities (Murray & Lopez, 1997). The growing impact of chronic disease, mental disability, and injury also contribute to this burden (Yach et al., 2004), as do the risks associated with globalization and harmful lifestyles, for example, pandemics, tobacco use and drug use. Frenk (2009) describes this trend as the “triple burden of disease”.

The city of Ahmedabad in Gujarat state has a population of 6.3 million, including 2.5 million who live in slum or slum-like areas (38% of total population). This is in comparison to the slum rate of 52% in New Delhi and 54% in Mumbai. The health divide along socio-economic status in Ahmedabad reflects the national trend (Somani, 2012) and serves as an appropriate case study for urban health inequity and waterborne disease burden. Diseases such as acute gastroenteritis, typhoid, viral hepatitis, and cholera are endemic to urban cities in India and disproportionately affect informal settlement areas where water resource infrastructure remains a critical problem (Agarwal, 2011). Slum areas are characterized by overcrowding, deterioration, and unsanitary conditions, making residents especially vulnerable to disease (UN-Habitat, 2004). The lack of adequate infrastructure comes from rapid urban growth, often unregulated; the Ahmedabad Municipal Corporation (AMC) is generally unable to provide proper construction and repair services with such high demand (Saravanan, 2013). The result is open-sewage and widespread illegal water connections, often of poor quality, which cause the mixing of sewage and drinking water. These conditions are a considered to be the primary cause of frequent outbreaks of viral hepatitis (hepatitis A and E) diseases centered in slum areas (Arankalle, 2012; Saravanan, 2013). Flood events or contaminated food served at restaurants and festivals are also an important source of viral hepatitis (Arankalle, 2012; Heathcote et al., 2004; Tandon et al., 1985).

Viral hepatitis diseases are pervasive waterborne, communicable, and infectious, disproportionately affecting the poor (Purcell & Emerson, 2008) and are endemic to India, including Ahmedabad. Every year, globally, 1.4 million people are diagnosed with hepatitis A and 20 million people with hepatitis E. The viruses are indistinguishable without a diagnostic blood test. They are spread through ingestion of contaminated water or food through the fecal-oral route, mostly attributed to poor hygiene and sanitation measures. Common symptoms include fever, loss of appetite, nausea, abdominal discomfort, and jaundice. People over the age of 12 are more likely to catch the virus compared to children, and there is no bias by gender except pregnant women are more vulnerable once they have contracted the disease. The incubation period is 14-21 days for the A strain and about 40 days for E, and a full recovery commonly takes months. The mortality rate is between 1 and 3 percent, but rises to 25 percent for pregnant women with hepatitis E

(Heathcote et al., 2004) Currently, there is no medical treatment available. Immunizations are available and effective, but, due to cost and access, often the most vulnerable populations go un-treated (Purcell & Emerson, 2008; Tandon et al., 1985). South Asia, and India in particular, bear the burden of these diseases: up to 60 percent of cases and 65 percent of fatalities occur in India.

For the 38 percent of Ahmedabad residents living in slum areas, access to health care is extremely limited. Ahmedabad only has 60 public health facilities, which can provide free or subsidized care. This is in addition to a general deficit in doctors, nurses, and other health workers nationally (Garg et al., 2012; Rao, 2005). For slum residents in Indian cities, the long distances required to travel to public clinics and costs incurred for seeking medical attention further restrict access to effective health care (Ensor & Cooper, 2004; Somani, 2012). A resource scarce environment, combined with the low quality of living conditions, directly contributes to disease risk. There is already little cash available for subsistence, leaving almost none for health care (Rao, 2005). Poverty is also a main driving force for tampering with water infrastructure (Vyas et al., 2010). Patil et al. (2002) reiterate this, linking already restricted livelihood opportunities with unsanitary environments (e.g., open sewage systems), further increasing risk. One common measure taken by private households for safe drinking water is the use of in-home water purification systems. Costs for the device and regular maintenance are high. Even then, most purification systems available on the Indian market are unable to filter the hepatitis E virus (Verma & Arankalle, 2009).

2.2.3 Possible explanations for the health gap

Scholars point to two important mechanisms in which this health gap exists and continues to grow in India: an underperforming and underutilized public health system (Binnendijk et al., 2012;

Ensor & Cooper, 2004; Patil et al., 2002), and an unsuccessful government supported drive for the commercialization of health care (Balarajan et al., 2011; Patil et al., 2002). Inspired by the Alma Ata declaration, India developed its first official National Health Policy in 1983 (Balarajan et al., 2011). An emphasis was placed on primary health care as well as decentralization of the system (India, 1983). Yet, the gap in health outcomes for urban and rural residents, and scheduled castes[1] remains high (Patil et al., 2002; A. Singh et al., 2011). Desired health outcomes that have not yet been achieved include: increased life expectancy, increased vaccination rates, improved access to health resources for the poor, and decreased child and maternal mortality, and decreased malnutrition rates (Joumard & Kumar, 2015).

Duggal and Gangolli (2005) posit that the national health policy of India was adopted merely as a quick fix with an overly biomedical approach and failed to address the socio-economic and political environments underlying public health challenges. The authors (*ibid*.) argue that one failure lay in the government's adoption of a vertically oriented disease control program approach because it was popular with the major funding organizations and did not require significant investment in comprehensive health networks. A consequence of this vertical approach is that health education has been neglected and there is a high dependence on care from unqualified health practitioners, or "quacks" whose services are commonplace throughout slum areas and who capitalize on limited regulation (*ibid.*). Such health practitioners also include limited opening hours, high absenteeism amongst health workers, and low-quality medical supplies that reduce the overall capacity of public health services. As a result, the health gap has widened, further undermined by neoliberal policies, including privatization, beginning in the 1980s which cut investments in social services to further increase the health burden on the urban poor (U. Patnaik, 2007; Rao, 2005).

---

[1] Scheduled castes are historically disadvantaged groups recognized by the Indian government for their low socio-economic status.

To fill the void in government expenditure on health, private care has been heavily promoted since the Structural Adjustment Programs (SAPs) of the 1990s (Ghosh, 1996). The private sector has not been sufficient to serve the needs of the poor. The out of pocket cost falls heavily on already cash-poor households (Balarajan et al., 2011; Duggal & Gangolli, 2005). There also is a private sector bias towards curative services in lieu-of preventive or primary care, which cluster in richer urban centres (Balarajan et al., 2011). This combination of poor public investment in health care, a push towards privatization, and an underperforming private sector warrant the investigation into improved and novel health interventions, especially for controlling outbreaks for diseases like viral hepatitis (Chauhan et al., 2010; Vyas et al., 2010).

2.2.4 mHealth: possibilities and limitations

The term mHealth, coined by Robert Istepanian in the early 2000s, refers to the "exploitation of the mobile telecommunication and multimedia technologies and their integration into new health care delivery systems" (Istepanian et al., 2004; Istepanian & Lacal, 2003; Istepanian et al., 2006). The 'm' comes from mobile. Mobile phones are thought to possess flexibility, communicative functions, and connectivity, which can be used for disease prevention, surveillance and management, and treatment compliance (Déglise et al., 2012). (Déglise et al., 2012). In their framework for mHealth, Labrique et al. (2013) describe twelve common applications of mHealth, including: client education, point-of-care diagnostics, event tracking, data collection, health records, decision support, communication between patients and health providers or health workers, scheduling and planning, provider training, human resource management, supply chain management, and financial transactions. The potential for mHealth in developing areas, such as India, is attributed to the rapid and widespread diffusion of mobile phones (DeSouza et al., 2014;

P. N. Mechael, 2009). By 2002, there were more mobile phones than landlines in developing countries; by 2017 it is predicted there will be more mobile phones than people on the planet (Beratarrechea et al., 2013).

Facilitated through government support and strong competition, the mobile phone industry in India is growing, with prices decreasing and infrastructure being readily available in many areas (Gupta & Jain, 2012; S. K. Singh, 2008). Given the burden of waterborne disease in developing areas and dynamic nature of risk factors through contested water infrastructure, mHealth needs to be tested for its potential to address outbreaks like viral hepatitis. This claim is reiterated in the literature on health and development (Ensor & Cooper, 2004; Frenk, 2009; Jamison et al., 2006), as well in discussions on the potential for mHealth (Agarwal & Lau, 2010; Beratarrechea et al., 2014; DeSouza et al., 2014; A. Singh et al., 2011; Yadav et al., 2010), especially in India. This thesis uses the term 'disease surveillance', defined by Langmuir (1963) as the process of data collection, analysis, and dissemination of results for the control of disease outbreaks.

The specific scope and scale of mHealth applications are varied. A. Singh et al. (2011) propose a framework for rural health workers to collect health data via a web-enabled mobile phone connected to a central database using web-based applications. Medical professionals then analyze the health data and make individual recommendations for patients by communicating with them over phone. Haberer et al. (2010) report on the challenges and opportunities of short message service (SMS)[2] and interactive voice recording (IVR) for Antiretroviral Therapy adherence. They identify several barriers: cost, motivation, and training, but also highlight the success of the technology and its supporting infrastructure. They conclude such a system is feasible for a resource-limited setting. Many mHealth applications track infectious diseases. mHealth can help

[2] SMS is the most common mode of text interaction on a mobile phone. It consists of short text messages exchanged between mobile and/or fixed line phones using standardized communication protocols.

address chronic health issues (Agarwal & Lau, 2010; Breslauer et al., 2009; Déglise et al., 2012). Agarwal and Lau (2010) examine how medical sensors, for example, a glucometer or a magnifying glass with a camera, can be connected to mobile phones through Bluetooth to monitor symptoms and record data. An example of a broad approach to health care is the Dristhi Smart Registry (Mehl et al., 2014). Dristhi is a comprehensive monitoring system that was designed to strengthen the ability for health workers to deliver women's health services in rural populations by replacing paper-based patient records with mobile phone data, complemented with the use of decision support systems.

Researchers argue four prime challenges facing public health in developing areas can be addressed through mHealth. First, the cost of service delivery may be reduced through fast and reliable information access, for example, community health workers accessing patient records, or conversely, patients using telemedicine hotlines (DeSouza et al., 2014; Krishna et al., 2009). Second, health education can be facilitated through promoting good health practices via SMS or calling campaigns, downloadable applications, or web-enabled mobile content (Déglise et al., 2012). Third, in theory, health monitoring through mobile data collection methods can reduce the spread of disease and increase the efficiency of response (DeSouza et al., 2014; A. Singh et al., 2011). Finally, the compatibility of mobile phones with external devices can reduce the need for travelling long distances for certain medical device readings (e.g., blood pressure) and check-ups (Yadav et al., 2010). Owing to the relatively recent emergence of mHealth, these functions require further research to validate their application in the context of public health.

One important critique on the use of mHealth is the lack of reliable evidence and case studies demonstrating its long-term use or tangible success (Estrin & Sim, 2010; P. N. Mechael, 2009; Tomlinson et al., 2013). To date, most of the literature and grey literature on mHealth describe individual and small-scale projects using mobile phones in various capacities (Estrin & Sim, 2010). The lack of substantive theory and best-practice creates an appeal without direction. Where mHealth is used in large-scale applications, it is often the case that the “m” or mobile component is relegated to a periphery function in the overall program (P. N. Mechael, 2009). This critique questions the actual and/or potential role of the mobile technology.

Limited mHealth theory means technical and policy coordination for scaling-up applications by health agencies is complicated without clear direction to leverage or evaluate mHealth (Tomlinson et al., 2013). The lack of metrics available make it difficult to quantify the actual health outcomes related to mHealth tools and provide little direction for system evaluation (P. N. Mechael, 2009). In the future, frameworks for evaluation of mHealth can help structure long-term plans and policy for the scaling of mHealth approaches.

Secondly, for developing countries, a major barrier to the use of mHealth is the level of technology and technical proficiency required for effective engagement. This critique is related to the Digital Divide, or technological inequality in terms of access, proficiency, and usage opportunities, which is pertinent for developing countries (Norris, 2001; S. Singh, 2012). S. Patnaik et al. (2009), Mehl et al. (2014), and Déglise et al. (2012) identify limited technical expertise and the need for extensive training, for both health workers and patients, as limiting factors in the deployment of mHealth. The nature of mHealth, with ongoing innovation, also means it can be difficult for organizations to keep their systems and architecture up-to-date (Estrin & Sim, 2010). Because of the degree to which private industry contributes to advances in mHealth, public-private partnerships are proposed to minimize this consequence (P. N. Mechael, 2009). There are, however, potential risks related to such partnerships, including, a clash of priorities (e.g., profit vs. health), and proprietary data and technology arrangements that complicate data confidentiality

(Barlow et al., 2013; Kraak et al., 2012). These critiques contribute to the difficulty in establishing direction in mHealth protocols.

Scholars also point towards the need for increasing geo-spatial functionality within mHealth. Fiordelli et al. (2013) comment that a majority of mHealth initiatives do not capitalize on the additional benefits of smartphones (e.g., GPS and Internet connectivity). As such, the full potential for spatial applications with mHealth, specifically in India, have been underdeveloped. Labrique et al. (2013) highlight the importance and relative theoretical ease, yet lack of activity, in exploiting location based services (LBS) on mobile phones for performing spatial analysis of health data. A. Singh et al. (2011) also reiterate the need for integrating spatial analysis into mHealth disease surveillance tools to replace paper records. There have also been studies advocating the need for near real-time geo-surveillance, especially in emergency situations (Butler, 2006; Carneiro & Mylonakis, 2009; Nsubuga et al., 2006; van Dijk et al., 2008). mHealth can integrate spatial analysis methods because it allows for collection of location data and storage in a spatial database. The effectiveness of this functionality, however, has rarely been validated.

2.3 Conclusion

Waterborne disease is influenced by geography. The ability to identify a spatial hot spot is one indicator of sensitivity for surveillance and the main objective for geo-surveillance. This makes it the ideal case in point for capturing the spatial location of individual occurrences within surveillance systems, which can then be detected as hot spots using post-hoc methods. By identifying the location of hot spots for disease in underserved or overburdened populations, public health officials can target limited resources to those areas with appropriate public health interventions.

This literature review demonstrates the need to examine public health applications designed to address waterborne disease. mHealth has been proposed as a potential tool in resource-poor settings, yet there has been little research to validate its application and success, especially when leveraging geo-spatial functionality. This is the research gap my thesis aims to contribute towards. First, by examining the existing surveillance system, it will be possible to see if and where challenges exist in Ahmedabad for monitoring the outbreak of waterborne disease. I will then explore the possibility of approaching surveillance with an open source and geo-spatial mHealth tool. This research will provide a case-study that can be used to evaluate calls in the literature to increase the use of spatial methods with mHealth applications.

Chapter 3. Critical evaluation of disease surveillance in Ahmedabad

3.1 Role of Manuscript in the Thesis

The role of this chapter in the thesis is to evaluate a specific surveillance program, the Integrated Disease Surveillance Programme (IDSP) in Ahmedabad, using a framework based on evaluation guidelines proposed by the US Centre for Disease Control (CDC) and World Health Organization (WHO) for disease surveillance systems. This chapter thus responds to Objective 1 of this thesis. This evaluation is limited to waterborne diseases in Ahmedabad for the purposes of this study and focuses on the ability to carry out geo-spatial disease surveillance. Although the scope of this evaluation is at the local level, the national IDSP framework will briefly be mentioned to provide context for the local deployment. This chapter highlights several key findings that suggest the need for investigating alternative approaches to disease surveillance.

3.2 Abstract

The city of Ahmedabad experiences frequent and widespread outbreaks of waterborne disease, often attributed to illegal and undocumented tampering and repair of the city's water pipe infrastructure. Evaluation of the strengths and weaknesses of the existing disease surveillance systems in Ahmedabad is a prerequisite to identifying potential systematic changes that can be made to improve upon the function of the system. This chapter examines the role of the Integrated Disease Surveillance Programme (IDSP) in Ahmedabad for carrying out effective waterborne disease surveillance, with a focus on geo-spatial disease surveillance. I assess the following system attributes: simplicity, flexibility, data quality, acceptability, sensitivity, predictive positive value, representativeness, timeliness, and stability. Overall, I found the simple design of the IDSP

and use of guidelines and protocols are its key strengths, however, data quality and timeliness pose challenges to carrying out effective waterborne disease surveillance. In its current format, the IDSP cannot be used for rapid and accurate geo-spatial disease surveillance for the identification of disease clusters.

3.3 Introduction

The impact of disease on livelihoods and health can be high in areas without adequate control or response mechanisms to disease outbreaks. A lack of public health resources in developing areas, including limited reporting mechanisms, a lack of intervention resources, and shortages in health personnel can often lead to undetected or uncontained outbreaks (Mehl et al., 2014). Additionally, individuals in low income countries with a high disease burden often face a lack of personal resources to help cope with disease, including frequent and ongoing outbreaks. Active disease surveillance may help reduce this burden by accurately identifying outbreaks during their early stage which can trigger response mechanisms to control their spread.

The city of Ahmedabad faces frequent and widespread outbreaks of waterborne disease. Diseases such as acute gastroenteritis, typhoid, viral hepatitis, and cholera are particularly pervasive, especially in informal settlement areas. These outbreaks are often attributed to poor water resource infrastructure, compounded with overcrowding and a lack of maintenance (Agarwal, 2011). The rapid rate of urbanization, especially informal settlement, means open-sewage and illegal water connections are common, and can often be directly linked to the contamination of drinking water (Saravanan, 2013). The magnitude of waterborne disease (number of outbreaks and affected population), rate of unregulated urban development, lack of public medical providers subject to official reporting requirements, and a personnel shortage within the

Ahmedabad Municipal Corporation (AMC) make disease surveillance difficult in Ahmedabad (Chauhan et al., 2010; Herma, 2013; Saravanan, 2013).

Monitoring high-risk areas for disease can help accelerate outbreak interventions and timely response (Morse, 2012). Given the nature of waterborne disease in Ahmedabad, often spread through unmonitored water infrastructure construction and repair, a proactive approach to surveillance may be suitable for resource allocation during a waterborne disease outbreak. Knowing the location of individual cases can be used to determine if and where an isolated outbreak of disease exists within a pre-defined area (Unkel et al., 2012). This is known as hot spot analysis and requires that cases are reported with geographic information, for example the city ward or street of the patient's home. Identifying the location of an outbreak of disease can potentially be used to help locate the risk factor associated with the outbreak and guide administrators in the design and implementation of a targeted response.

The IDSP is the national disease surveillance system of India. It is implemented at the national, state, and district (or city) level. The overall aim for the IDSP is to provide a decentralized system of disease surveillance for timely and effective public health action implemented at a relevant scale, in this case, the city level (World Bank, 2012a). It also aims to improve the efficiency of disease surveillance for use in health planning, management and evaluating control strategies. Specifically, the IDSP seeks to detect disease outbreaks as they emerge, and enact appropriate response mechanisms for disease containment.

Disease surveillance systems should undergo ongoing evaluations to improve reporting mechanisms, technical capacity, and the ability to adapt to current and changing public health needs. Evaluations can help guide the development of recommendations for system changes and should occur while the system is being used to understand existing strengths and resolve system challenges. The US' Center for Disease Control (CDC) and WHO recommend that system evaluations consist of an assessment of key system attributes: simplicity, flexibility, data quality, acceptability, sensitivity, predictive positive value (the ratio of identified cases to actual cases under surveillance), representativeness, timeliness, and stability (German et al., 2001; World Health Organization, 2006). It is thought that such evaluations can help improve disease surveillance activities and reduce the burden of disease (Alwan et al., 2010; Berkelman et al., 1994; Butler, 2006).

In this chapter, I present an evaluation of the IDSP in Ahmedabad. I use a review of grey literature, and a dataset of four months of case reports collected by community health workers and link workers under the IDSP. These data provide evidence to assess the performance of the IDSP for carrying out geo-spatial disease surveillance. The core system attributes that are evaluated include: simplicity, flexibility, data quality, acceptability, sensitivity, positive predictive value, representativeness, timeliness, and stability. I found that data quality and timeliness are two key system attributes that inhibit the overall ability to fulfill the local public health goals, especially as they relate to geo-spatial disease surveillance of waterborne disease.

3.4 Methodology

3.4.1 The Integrated Disease Surveillance Programme in Ahmedabad

The IDSP in Ahmedabad was set up in 2004 when it was deployed under the national framework through funding from the World Bank. The IDSP was proposed in response to limitations of a prior program, the National Surveillance for Programme Communicable Diseases (NSPCD). Broadly, the programme consists of case based reporting from surveillance of selected priority diseases (Table 1) with consequent analysis and interventions implemented at the appropriate scale, for example, if an outbreak crosses state lines (National Informatics Centre, 2009). The National Informatics Centre (NIC) (2010) provide standard case definitions, formats, and training for individual disease case reporting, intended to ensure uniform surveillance and analysis methods throughout the reporting units. The city of Ahmedabad is required to adhere to these protocols to ensure uniform reporting measures when dealing with inter-regional issues.

Table 1. Priority diseases monitored under the IDSP (National Informatics Centre, 2010)

Name of Disease		
Enteric Fever	Diphtheria	Pneumonia
Dengue Fever	Cholera	Acute Respiratory Disease
Viral Hepatitis	Bacillary Dysentery	Acute Flaccid Paralysis
Leptospirosis	Encephalitis	Malaria
Measles	Pertussis	Chikungunya
Meningitis	Chicken Pox	

At the local level, case reports are submitted to a central district surveillance unit (DSU) by public health workers assigned to specific sub-localities. The DSU then aggregates new cases by disease and date. If an outbreak is suspected, the DSU will submit a report indicating the context and nature of the outbreak in a web portal. Next, a Rapid Response Team (RRT) is sent on the recommendation of the public health authority to investigate and act upon the possible outbreak.

3.4.2 Evaluation Framework

Effective disease surveillance requires that each step in the process is followed meticulously and efficiently, otherwise the system will break down, limiting overall effectiveness (German et al., 2001). Limitations in health systems will inevitably hinder health outcomes in the target populations for such systems. Berkelman et al. (1994) highlight the need for inter-agency cooperation, prompt diagnosis and reporting mechanisms, technical capacity able to respond to current and changing public health needs, and regulatory support. Thacker et al. (1983) report that a lack of representativeness, timeliness, sensitivity, and/or specificity of case reports are the most critical limitations for traditional disease surveillance.

Appropriate and useful system evaluations are designed to guide the development of recommendations about the function of the system. To thoroughly implement, and subsequently develop disease surveillance systems, evaluations should be carried out during a deployment. This will improve the quality, efficiency, usefulness, and outcomes of the system, and in turn help identify and solve systematic limitations. The US' Center for Disease Control (CDC) propose that an evaluation consists of an assessment of the following key system attributes: simplicity, flexibility, data quality, acceptability, sensitivity, predictive positive value (the ratio of identified cases to actual cases under surveillance), representativeness, timeliness, and stability (German et al., 2001). The WHO has a similar approach (2006), with one distinct variation: the differentiation and importance of both monitoring and evaluation. The inclusion of monitoring is complementary to a system's evaluation. Monitoring involves the continuous tracking in all phases of implementation to ensure the specific system targets are being met. Monitoring can ensure the best use of resources by reallocating them depending on ongoing outcomes and indicators. It is an ongoing process, ideally carried out separately from evaluation. While the WHO's recommendation for monitoring is pertinent and important; however, monitoring is outside the scope of this research because of its specific focus on taking immediate action, tracking progress, and updating system plans.

I used the principal system attributes proposed by the CDC for evaluation of disease surveillance systems to create an evaluation framework on the ability for the IDSP to carryout effective disease surveillance. I focus the evaluation on the geo-spatial aspects of disease surveillance. This evaluation consists of an analysis of the system attributes (Table 2).

Table 2. System attributes to evaluate geo-spatial surveillance process

System Attribute	Definition (From - German et al., 2001)	Evaluation criteria
Simplicity	The simplicity of a surveillance system refers to both its structure and ease of operation. Systems should be as simple as possible while still meeting their objectives.	Level of integration across components; use of operational procedures.
Flexibility	A flexible surveillance system can adapt to changing information needs or operating conditions with little additional time, personnel, or allocated funds. Flexible systems can accommodate, for example, new health-related events, changes in case definitions or technology, and variations in funding or reporting sources.	Ability to adapt to changing system needs
Data quality	Data quality reflects the completeness and validity of the data recorded in the public health surveillance system.	Rate of missing, ambiguous, or unknown data in case reports
Acceptability	Acceptability reflects the willingness of persons and organizations to participate in the surveillance system.	Stakeholder engagement
Sensitivity	The sensitivity of a surveillance system refers to the ability to detect outbreaks, including the proportion of cases that are captured and identified as cases within the system that can be used for cluster detection.	Ability to conduct hot spot analysis (LISA)
Positive Predictive Value (PPV)	Positive Predictive Value (PPV) is the proportion of reported cases that actually have the health-related event under surveillance	Outside scope of study
Representative ness	A surveillance system that is representative accurately describes the occurrence of a health-related event over time and its distribution in the population by place and person.	Outside scope of study
Timeliness	Timeliness reflects the speed between steps in a surveillance system.	Length of reporting delays and differences in

		average delays across city wards
Stability	Stability refers to the reliability (i.e., the ability to collect, manage, and provide data properly without failure) and availability (the ability to be operational when it is needed) of the surveillance system.	Obstacles to regular reporting mechanisms including data quality and timeliness, usable reports and reporting delays

3.4.3 Analysis

The system attributes being evaluated measure the effectiveness of the performance of the system for data collection and analysis. Based on reliable, valid, and informative evidence on these attributes, the ability of IDSP to monitor waterborne diseases in Ahmedabad was assessed. This was done using grey literature obtained from government and NGO websites, system documents (i.e., training manual and funding assessments), and a dataset of four months' worth of case reports collected in normal operations for the IDSP provided from the AMC Health Department to evaluate the system. This dataset contains case reports collected between April and July 2014 (2,194 entries) by community health workers and hospital staff attributed to local and sub-district reporting units in Ahmedabad for four highly contagious diseases (i.e., viral hepatitis, cholera, acute gastroenteritis, and typhoid). Each case report contains information such as the patient's home address and ward, diagnosis type, diagnosis date, and diagnosis source. Table 2 provides a definition of each system attribute to be evaluated and how it was operationalized for this analysis.

Simplicity was qualitatively evaluated on the basis of the operational design and protocols established for the IDSP. The simplicity of an effective surveillance system should ensure procedures are clearly defined and documented to ensure all stakeholders are fully engaged, and allow for inter-agency cooperation through the use of clear and standardized methods.

Flexibility within a surveillance system allows for it to adapt and incorporate changing health priorities during its operation. At the national level, the IDSP should allow individual state and district units to monitor regionally important diseases. This enables locally relevant deployments of the IDSP that can target local public health concerns. The effectiveness and ease of the IDSP at the local level to expand its surveillance activities to monitor priority diseases as they emerge is a measure of its flexibility.

The validity of a surveillance system is directly related to quality of data collected through case reporting. This means that data collection methods must ensure consistently high levels of data accuracy and precision, as well as using standardized case reports with a consistent typology. Data quality can influence the strength of analysis and ability for the system to confidently meet its intended purposes by identifying true outbreaks of disease. It is critical to adhere to the established case reporting standards defined through the operational framework (Sahal et al., 2009). I used 2,194 case reports from the IDSP in Ahmedabad to determine the rate of unusable data as a measure of data quality. Unusable data is defined as missing, duplicate, or incorrect values submitted in a case report which cannot be used for accurate geo-spatial analysis of disease data.

The acceptability of surveillance system will ensure its long-term sustainability and continued user engagement. Acceptability is measured by the willingness of stakeholders to contribute to the IDSP. The participation of health workers responsible for case submission, the local government, and health providers in Ahmedabad are evaluated to measure acceptability. . Another measure of acceptability is growth in capacity for outbreak detection.

Sensitivity refers to the ability for the system to detect actual disease outbreaks. This depends on multiple attributes: data flow, including quality, and analytical methods used for outbreak detection. Sensitivity is especially critical with regards to waterborne and vaccinepreventable diseases. If a system cannot accurately detect an outbreak, a waterborne outbreak is likely to outdistance its treatment. Sensitivity in this study was evaluated as the ability to carry out geo-spatial disease outbreak detection through spatial statistical analysis of cases recorded in the system, specifically using Local Indicators of Spatial Analysis (LISA) and the GI* method, which are standard approaches in spatial epidemiological outbreak investigations (Kitron, 1998). These methods have been found to be effective in detecting outbreaks (Chen et al., 2011; Colin Robertson et al., 2010) and is relevant in the Ahmedabad context for waterborne disease. Outbreak detection is challenging when simply considering spikes in reports from reporting units because this fails to account for the spatial distribution of cases and local incidence of spatial risk factors.

From a public health perspective, timeliness relates to the delay from illness to intervention or follow up. In this evaluation, I measured one component of timeliness associated with data management: the delay in aggregating individual case reports. Delay is calculated as the number of days from initial diagnosis at a medical facility to the case report being included in the aggregated dataset. Another indicator of timeliness is the variation in delays across reporting units and by disease. I used the Kruskal-Wallis equality-of-populations test to evaluate the difference in reporting delays by disease and city ward. It should be noted that other components of timeliness not measured in this evaluation are related to delays in diagnosis and initial reporting.

Finally, the stability of the IDSP and its related activities was measured. The CDC recommends that a system be able to carry out effective surveillance during all times deemed vulnerable for the priority diseases. In practice, this implies that the IDSP be able to operate year-round due to the risk for priority diseases in Ahmedabad.

3.5 Results

3.5.1 Simplicity

The system follows a tiered organizational structure with reporting units attributed to each district surveillance unit presented in Figure 2. According to the IDSP training manual for health workers (National Informatics Centre, 2010), case reports are completed at the local or sub-district level (i.e., community health centre, sub-centre, and hospitals) and aggregated to district surveillance units (DSU). This structure is simple and follows a clear hierarchy and logical data flow. Analysis within each DSU is initially completed to identify outbreaks through finding anomalies in the disease rate. Notable results are then reported into the national IDSP web interface, if necessary. State and national officials can access the reports and intervene using a Rapid Response Team (RRT) when required.

This operational structure for reporting, analysis, and intervention using NIC standards adheres to disease definitions of the WHO (2014) and is in line with international system

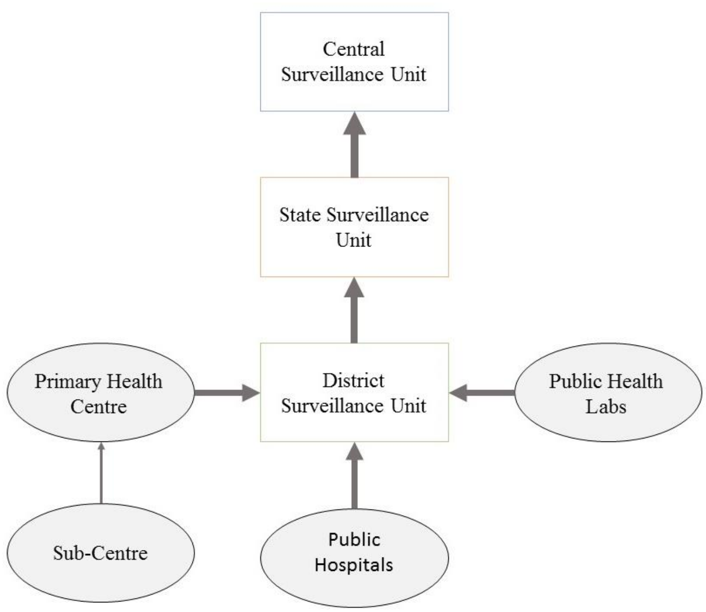

Figure 2. Operational structure of the IDSP

recommendations (Yasnoff et al., 2000). The standardized reporting mechanisms throughout the different reporting units, in theory, ensures inter-agency collaboration and integration. It is important to note that simplicity is primarily an attribute of the system's design, and not necessarily performance. Regular monitoring of the IDSP, as recommended by the WHO, may provide a more thorough overview of simplicity of the system.

3.5.2 Flexibility

Since the IDSP was first deployed, the number of priority diseases being monitored has increased. This reflects the relative and changing disease burden for urban Ahmedabad and flexibility of the IDSP for being able to monitor emerging diseases. An example is the inclusion of case reports for food poisoning, Lyme disease, and rubella in response to increasing public health burdens in Ahmedabad. At the national scale (thus, inclusive of Ahmedabad), the IDSP has responded to national emergency epidemics such as H1N1, avian flu, and mad-cow disease (World Bank, 2012b). The IDSP, through recognition of the importance of environmental factors, has expanded its scope to monitoring water and air quality as a risk factor for disease. It is unclear how local changes in surveillance priorities are integrated in terms of data definitions, clinical protocols, and analytical methods in an operational sense. Ideally, this evaluation would consider documentation of these mechanisms, however, documentation does not exist outlining these changes. At a high-level, the IDSP is flexible in its operations as measured by its ability to expand surveillance activities to changing public health priorities.

Flexibility is especially important in a developing area where health situations can change rapidly and priorities are often region dependent. The presence of infectious diseases in Ahmedabad is similarly dynamic. Globalization has meant that Ahmedabad is an increasingly connected city which increases exposure to global epidemics resulting in changing health priorities.

For example, in early 2015 the city experienced a major outbreak of swine flu. The ability for the IDSP to customize its surveillance activities at the local level is particularly useful for Ahmedabad.

3.5.3 Data Quality

Table 3 presents the results of the analysis on data quality. I found systematic errors in the case reports. Problem fields include: ward name, address or contact information, and age/gender/date formats. Up to 45 percent of all case reports were compromised through one of the error types in Table 3. Cases with erroneous ward values contained incorrect or inconsistent spellings of the ward name. For example, six cases reported in the Krushnagar ward (official AMC name) were recorded with the ward name value "Krushna nagar", five with "Krushnanagar", and one with "Krishnanagar". It is also troubling that the value "jaundice" is recorded for disease type instead of viral hepatitis or a specific type of hepatitis (i.e., A or E) as is recommended in the NIC's guidelines. Automated geo-spatial analysis requires a standard typology and identical classification of data for accurate analysis. Data points with erroneous values or variations of the same spelling will be considered separate values in an automated geo-spatial analysis. Additionally, four percent of reports contained ambiguous values, such as imprecise address values like "SANTOSHINAGAR" and "Rabari colony" that would make it difficult to identify the patient if any potential follow ups by AMC staff were required.

Table 3. Evaluation of data quality from IDSP case reports in Ahmedabad April-July 2014

Type of Error	Evaluation
Ward values	The AMC divides the city into 64 city wards. 86 unique ward names were reported. A total of 419 reports (19%) contained erroneous ward values which cannot be used to accurately identify the actual ward of the patient.

Contact Information	568 reports (26%) did not contain a home address or phone number.
Multiple Entries	Three patients had three unique case reports filed, and 58 patients had two unique case reports filed for the same health event. A total of 64 reports (3%) are considered to be multiple entries.
Missing values	A total of 81 reports (4%) contained at least one missing value or an ambiguous value.
Total	A total of 581 unique reports (those without contact information - 27%) contained erroneous values or were unusable in analysis. When case reports without proper contact information are included, this number is 986 (45%).

My evaluation found that data quality of many case reports at the local level is compromised through one of the error types listed in Table 3. Twenty seven percent of unusable data far exceeds an acceptable error rate (10%) in an operational public health system (Arts et al., 2002). Such data quality suggests issues in data collection, reporting, and aggregation. In an operational sense, this limits the ability for accurate disease outbreak detection attributed to systematic uncertainty.

Errors in the location of disease cases from mislabelling ward names makes it difficult to perform spatial analysis. This compounds the locational inaccuracies, which in any case can only be located to a specific ward and not to GPS co-ordinates or home addresses. The occurrence of multiple entry case reports include duplications of the same information, entries with different name spellings, and reports entered on two different dates, but for the same health event. This indicates a lack of coordination in case reporting and/or aggregation. Missing or ambiguous field values can also inhibit strength of analysis. Data quality seems to be linked with the acceptability and simplicity attributes.

3.5.4 Acceptability

District, state, and national public health agencies have committed to future funding and participation in the IDSP because of perceived success, including the AMC. Since its inception in only nine states, the IDSP has been fully adopted across all 35 states. Further indication of the acceptability of IDSP was the establishment of a dedicated call centre to provide citizens with up-to-date disease outbreak information.

Since 2008, the number of outbreaks detected nationally has increased (Figure 3). This does not necessarily reflect the growing burden of disease, however, since establishing surveillance programs invariably increases reporting. This increase over time may be partly a proxy for the increased participation of different surveillance units that are capturing previously undetected outbreaks (National Informatics Centre, 2009). In their final funding evaluation, the World Bank stated that it was satisfied with the national deployment of the IDSP because it has contributed towards the establishment of new public health networks and engagement at relevant levels (i.e., across each state and within pertinent cities/districts) (World Bank, 2012b).

Locally, one specific point of contention in the acceptability of IDSP in Ahmedabad is disagreement between local link workers and the municipal government over wages and contracts, resulting in ongoing strikes ("Link workers' protest against Ahmedabad Municipal Corporation," 2014). Anecdotal evidence also suggests it is common for link workers to delay submitting case reports (D. Saxena, personal communication, June 3, 2014). Although it is difficult to measure within this study, it is widely accepted that many case reports are never actually submitted in the first place (*ibid*). Furthermore, though not directly mandated through IDSP, private labs and health care providers are not currently mandated to provide case reports to the DSU.

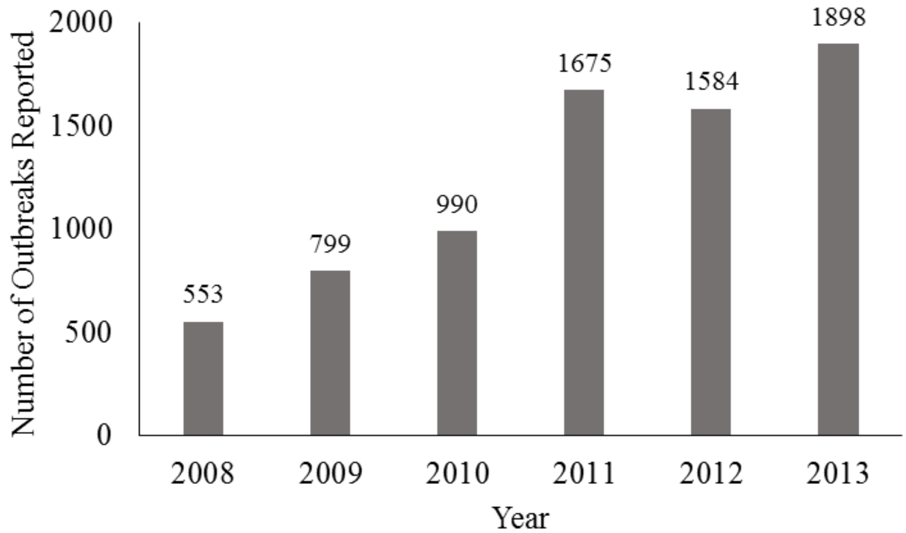

Figure 3. Number of outbreaks detected nationally under IDSP

The willingness for stakeholders at each level within the system to engage in the IDSP framework is an indicator of cohesion and operational acceptability. Stakeholders include persons and organizations directly engaged in the system, including patients who provide data, data collectors, and persons directly affiliated with the public health agency. Stakeholders must be willing and capable in their respective roles to facilitate the overarching public health goals. At the national level, the acceptability of IDSP appears high. Continued willing participation in the program from the city of Ahmedabad indicates the city has benefited or sees potential in the continued use of the program. Conversely, it may suggest no valid alternative.

Link workers play a critical role at the frontline of disease surveillance in Ahmedabad, including submitting case reports. Any disruptions to their IDSP engagement will detract from overall operations. Because of the ongoing labour disputes, data collection may be compromised. These disruptions have been ongoing for several years and may be linked to delays in reporting and issues with data quality. Interviews with link workers may be useful for exploring this point further. A second major barrier to general acceptance of the IDSP is the lack of participation from private providers. In its current design, the IDSP makes no mandate for this inclusion, however, in the context of public health, these are important stakeholders, whose lack of commitment damages the ability to ensure broad acceptance.

3.5.5 Sensitivity

Although most case reports include a home address, it is essentially impossible to geocode these values due to the subjective address structure recorded by data collectors (e.g., Y address is across from X landmark) without an appropriate georeferenced base (which does not exist). This is further compounded by the growing prevalence of informal housing, often without an actual government address. Thus, cases cannot be plotted spatially unless a large areal unit is utilized (e.g., a municipal "ward" as opposed to a specific house address). In such cases, the data do not provide a sense of "closeness" for each observation, making it difficult to use spatial outbreak detection methods which rely on exact locations for their calculations.

The most reliable spatial information collected in case reports are the wards of each patient. Therefore, hot spot analysis can only be run on aggregate case counts at the ward level. In Ahmedabad, only aggregate count of cases by Ward are used for evaluation, ignoring other spatially explicit methods to detect clusters of emergent diseases. Here, the GI* method is tested as the mechanism by which wards with a higher than expected disease rate (based on the general disease rate) can be identified (Figure 4).

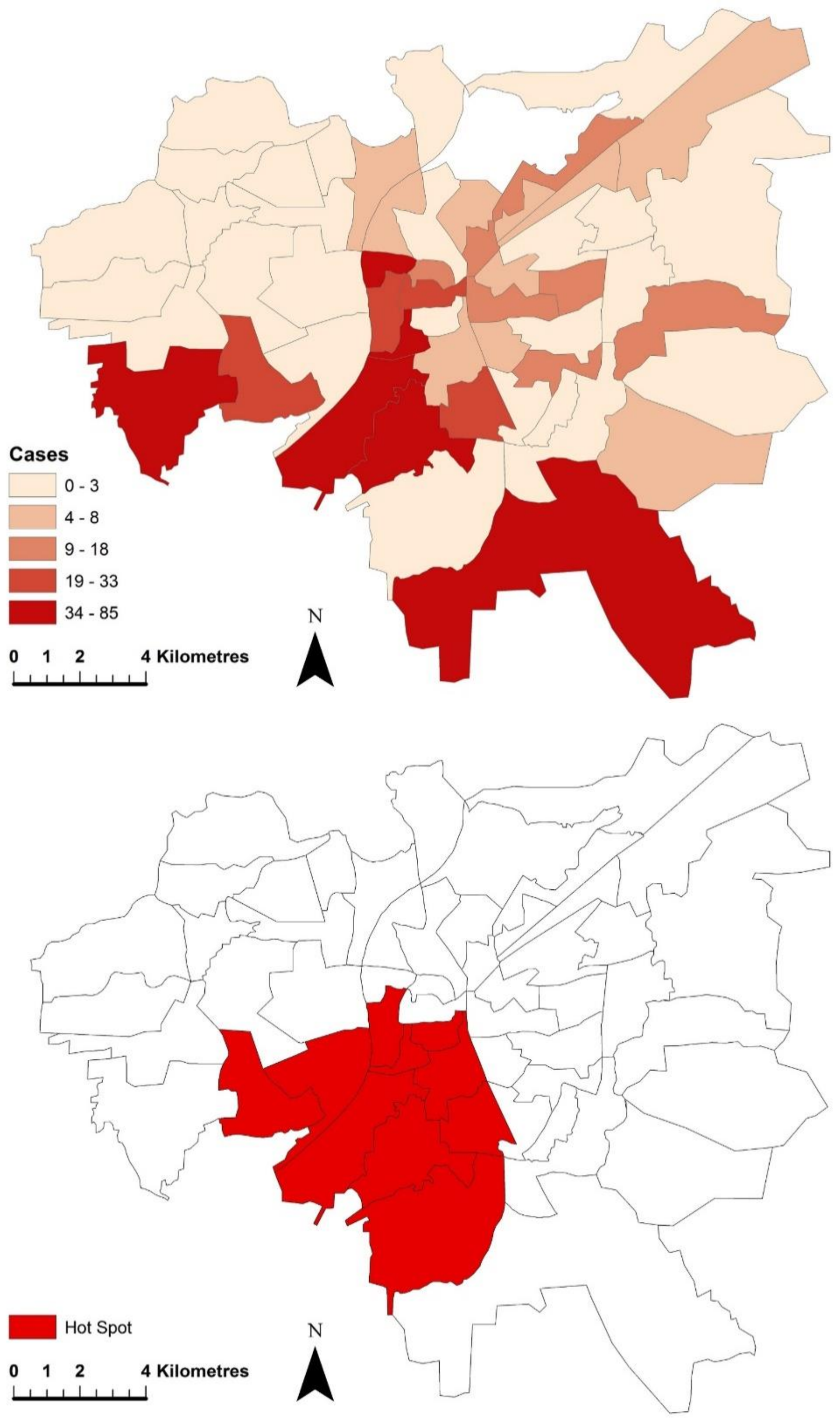

Figure 4. Example of hot spot analysis capacity under current IDSP framework

The analysis of location data quality suggests the ability to perform hot spot analysis is compromised without having a measure of distance between cases. One could remove the need for address reporting with, for example, GPS coordinates to provide a sense of “closeness” for disease incidence. In the absence of GPS coordinates, hot spot analysis can be done at an aggregated level, for example, city ward or zone level. The coarse precision of analysis, however, may limit the effectiveness of any measures to contain the outbreak at its source.

3.5.6 Positive Predictive Value and Representativeness

The positive predictive value (PPV) refers to the likelihood that an event labelled as an outbreak is actually an outbreak, and is related to sensitivity. Reporting a false positive as an outbreak may result in unnecessary use of resources and detract from overall public health activities. Representativeness is similar in that it assesses how well case reports are able to capture the overall disease situation present in the reporting unit. Though an important aspect of disease surveillance, PPV is difficult to evaluate without an alternate source of case reports (which was not available in the case of IDSP).

3.5.7 Timeliness

A breakdown of delays of case reports, which were not compromised in terms of data quality (excluding missing contact information), are presented in Table 4-6. Frequent and lengthy delays of case reports by disease type and ward were identified in these data. It was also noted that aggregate delays in reporting differed across city wards, which may indicate non-systematic differences in data collection and reporting mechanisms throughout the city. If a speedy data flow of case reports can be ensured from the field to the DSU and then to the RRT, the system can ensure that a response is implemented quickly to reduce the impact of a detected outbreak by preventing further cases. Due to the nature of the target diseases, delays in any of these steps will have direct impacts on health outcomes (Jajosky & Groseclose, 2004).

Table 4. Overall reporting delay for uncompromised case reports

Reporting Delay	Number of cases	Percentage of cases (out of 1613)
> 1 Day (average delay 3.9)	382	24
> 3 Days (average delay 6.2)	173	11

Table 5. Reporting delay for problematic city wards

Ward	Average Reporting Delay (days)	% Over 1 Day	% Over 3 Days	N (Number of case reports)
Meghaninagar	4.4	84	56	25
Naroda	4.3	72	41	22
Odhav	2.1	36	15	39
Bapunagar	2.0	15	15	97
Vastral	1.7	38	5	21
Kalupur	1.7	30	14	57

Notes: The Kruskal-Wallis equality-of-populations test confirms each ward has a different mean with $p = 0.0001$.

Table 6. Reporting delay for case reports by disease

Disease	Average Reporting Delay (days)	% Over 1 Day	% Over 3 Days	N (Number of case reports)
AGE	1.3	19	4	918
Cholera	3.3	80	57	35

Jaundice	1.9	22	14	522
Typhoid	3.1	44	33	138

Notes: The case report delay distribution is lognormal. The Kruskal-Wallis equality-of-populations test confirms each disease has a different mean with p = 0.0001.

Within the past twenty years, many countries have adopted electronic reporting mechanisms which mandate or achieve a reporting delay for communicable diseases of one day or less, with success in terms of outbreak detection (Krause et al., 2007; Reijn et al., 2011). Electronic reporting is an integral part of the IDSP from the DSU level and higher. ICTs are not as readily available at the local level in labs, hospitals and health centres, making electronic reporting difficult from a diagnosing facility. The frequent and lengthy delays in getting a case report to the DSU are a challenge for conducting up-to-date analysis on a regular basis. Temporal clustering is also challenged by the slow and unpredictable rate at which case reports are submitted. Though it is possible to perform post-hoc temporal analysis, early outbreak detection is sub-optimal with the current reporting delays.

3.5.8 Stability

Delays in reporting, workforce disruptions, errors in case reports, a dependency on electronic analysis and data storage which is subject to frequent electricity cuts, and a paper-based reporting mechanism, bring into question the stability of the system to respond to emergency situations. For example, the stability of the IDSP is likely compromised during worker strikes, which limit the ability to collect case reports during those periods. Conversely, the direct integration of different levels of deployment of the IDSP (local, state, and national) is beneficial for surveillance of priority diseases in Ahmedabad which can be related to the health profile of surrounding regions. Another indication of the stability of the system is the integration of the surveillance and response mechanisms which may increase efficiency in the overall process.

For outbreak interventions to be effective at short notice, surveillance data and analysis must be reliable. The DSU is responsible for detecting outbreaks at the district level by looking for spikes in disease counts from each reporting unit (PHC, SC). If an outbreak is suspected, the DSU will submit a report indicating the context and nature of the outbreak to the web portal. Next, a Rapid Response Team (RRT) is sent on the recommendation of the public health authority to investigate and act upon the possible outbreak. Further investigation into the actual epidemiological methods being used within the AMC would benefit this discussion. Significant and frequent delays in case reporting reduce the ability for outbreak detection and make it difficult to implement rapid emergency interventions. The rate at which interventions can take place are mainly dependent on the rate at which reports are submitted. Data uncertainty and failure to adhere to established standards mean it is necessary to clean the data before any substantial analysis can be done. Additionally, in the disease field of case reports, the term "jaundice" is used instead of the NIC recommended 'viral hepatitis' or specific type of hepatitis (A or E). Such systematic reporting issues may further limit the stability of the system. Finally, the widespread workforce disruptions (as a result of the strikes initiated by link workers) potentially threaten the capacity for IDSP at the local level in the long-term.

3.6 Discussion: Evaluation summary and recommendations

This system evaluation provides evidence to suggest implementation of the IDSP at the local level is limited in its ability to carry out geo-spatial analysis to help achieve the stated public health goals. Two key attributes required for geo-spatial disease surveillance, data quality and timeliness, which are related to data collection and analysis, can be major barriers to surveillance (Berkelman et al., 1994). These attributes also relate to sensitivity of the system. Much of the disease data that do get collected are compromised under the current system. This stems from breakdowns throughout multiple procedural steps in the system. Initially, across the different reporting units there is minimal consistency in the way case data are reported. There appears to be no clearly defined data definitions, and if there are, definitions are not strictly adhered to. Further, location information (such as locational coordinates in latitude-longitude) for identifying cases of waterborne disease are not specified nor collected within the current IDSP framework.

Disease surveillance systems have to reliably and accurately aggregate case reports for analysis (Langmuir, 1963). Also highly problematic is the fact that the delays in reporting are statistically different across city wards and for different diseases. This is most likely due to macro-level factors, such as an unequal allocation of reporting resources across city wards, or intentional data manipulation amongst reporting units instead of individual error. Each of these issues is further compounded when the cases are eventually aggregated for analysis. With variations across wards and disease type in reporting delays, it is challenging to conduct any temporally sensitive analysis. For highly contagious diseases such as cholera or typhoid, failure to detect an outbreak at the same rate cases are being diagnosed unnecessarily places residents at risk for disease, and delays urgent and time sensitive interventions from a RRT. The ability to conduct timely analyses is also limited by the frequent need for manual data cleaning.

Sensitivity is evaluated with the Getis-Ord GI* analysis. The geo-spatial analyses that can be performed are limited as address locations are most reliable at the ward level. It is difficult to use point-based methods or visualization of cases due to the inability to geocode cases to specific locations. This limitation also means that any type of ward level analysis assumes an equal distribution of cases throughout the ward. This manifests itself in a mapping areal unit problem (MAUP) (Openshaw, 1984) where the size, shape, and relative borders of wards are ambiguous in terms of actual human settlement. Furthermore, ward level analysis makes no differentiation between land use and land cover, dwelling type, population density, or other environmental determinants related to disease. In practice, this means that designing and implementing a targeted intervention specific to an actual outbreak is difficult.

A limitation of this evaluation is the inability to conclusively assess positive predictive value (PPV) and representativeness. These attributes would be better evaluated through the use of regular monitoring of the system, as recommended by the WHO. The CDC also suggests calculating the number of false-positives reported to the IDSP through verifying diagnoses. PPV is an important measure to evaluate because false-positives can result in unnecessary depletion of scarce resources. Representativeness would most likely require data from another agency or surveillance system for comparison. Public-private partnerships may be one viable approach for this.

Based on the performance of the IDSP, Table 7 summarizes the evaluation of the IDSP.

Table 7. Summary of the evaluation of the IDSP by core steps

Attribute	Strengths	Weaknesses
Simplicity	Successful vertical integration	
	Clear system design and protocols	
	Allows for local level implementation	
Flexibility	Ability to incorporate long- and short-term health needs	
	Committed to disease surveillance in the long-term	
Data Quality		Poor adherence to standardized data formats and definitions

		Multiple errors, missing values, and duplications
Acceptability	Continued commitment from local, state, and national government	No provision for private provider engagement Labour disputes with link workers
Sensitivity	Provisions for suspected, probable, and confirmed cases	Difficult to carry out spatial analysis
Timeliness		Frequent and timely delays in reporting of cases
Stability	Standardized methods throughout national framework	Limited ability to invoke timely interventions

Although the IDSP is guided by a sound framework for disease surveillance, this evaluation concludes that the way it is currently implemented limits the capacity to achieve its given objectives. These limitations are of special concern for waterborne and other highly communicable diseases due to the dynamic nature of outbreaks and need for rapid and targeted interventions to contain risk exposure. To improve upon the existing system, the following recommendations are proposed: (1) adherence to standardized data formats; (2) improving the ability to provide near real-time case reports; (3) better integration of the private health sector; and (4) strengthening of spatio-temporal methods to detect outbreaks.

1) Adherence to standardized data formats

Acceptability and quality of data are intrinsically linked to the ability for data reporters to adhere to standardized reporting formats and data definitions. Therefore, to reduce the degree of unusable data being collected, a multi-step approach is recommended. First, the DSU should create a new set of locally relevant data definitions, allowable values, and data formats in line with international standards that are readily available to reporting units (World Health Organization, 2014). Panackal et al. (2002) note that such standardization ensures completeness of data across different reporting units for analysis at different scales. Second, all case reporters would be provided with comprehensive and regular training related to data input. Training ensures that disease information is collected according to the defined standards and properly submitted to the reporting unit (Arts et al., 2002). Next, the DSU should implement ongoing monitoring of data reporters through site visits and auditing, especially in problematic reporting units. It is important to consider, however, that in a low-resource setting, these steps may redirect resources from possibly more effective preventive measures or response efforts detracting from overall public health priorities. This thesis does not evaluate these trade-offs.

Finally, the implementation of electronic case reporting should be considered. Several systems already exist for this purpose including mobile phone-based solutions (mHealth), which compel the data collectors to use standardized values and formats. By following these recommendations, the information collected in case reports can be harmonized. This harmonization, in turn, I argue, will increase the strength of the overall system.

2) Ability to provide near real-time case reports

In any disease surveillance system, the timeliness of reporting is critical for its success. The frequent and lengthy delays, therefore, must be minimized. This will positively influence the ability to perform up-to-date analyses and ensure the containment of an outbreak before there are unnecessary new cases from a delayed response. The use of paper-based reporting is a possible barrier to this function because of the need for manual transcription and physical delivery of case reports. One suggested approach is the use of electronic reporting through the use of smartphones or other Information Communication Technologies (ICTs) to ensure the rapid and accurate collection of data, and subsequent aggregation of reports. These recommendations are in line with the disease surveillance systems of developed countries, and are being deployed throughout developing countries for similar reasons (Sahal et al., 2009). In practice, this would be quite challenging and needs to be approached with caution. Critiques of mHealth, including: a reliance on an often underperforming technical base, the need for ongoing training, lack of policy direction, unknown benefits to actual health outcomes, and difficulty in system evaluation are likely to be applicable in implementing such an approach in Ahmedabad and other cities in developing countries.

Due to the ubiquity of mobile phones in India, in addition to low operating cost and high competition levels, such a tool can be relatively cheap and easy to initialize (S. K. Singh, 2008). A mobile phone data collection system could be integrated into a free and secure web-server administered by the DSU for data aggregation and allow for close to real-time data updates. Multiple free and open source tools exist for mobile phone-based data collection including Open Data Kit, KoBo, and CommCare (P. Mechael et al., 2014). Data would need to be encrypted and protected through secure logins to ensure confidentiality (Panackal et al., 2002). An interface should be developed so the DSU would have full control over who is allowed to collect, submit, and access the data. The interface also would ensure reporting units are able to submit case reports without delay and directly from their own jurisdictions. Data security regimes would need to be established so data is stored safely for a long period of time while adhering to confidentiality requirements.

3) Better integration of the private health sector

A critical limitation in the IDSP is the inability to capture case reports from private health providers. This is critical because a majority of citizens use private services for their primary health care. This is largely in part due to limited government expenditure on, and capacity within, the provision of health care, the low cost of private care, and the large portion of informal health practitioners including spiritual healers or homeopathic doctors, which cater to the urban poor (R. Kumar et al., 2007). The ability for accurate geo-spatial outbreak detection is compromised through missing this large portion of data. Detection is further complicated by the segregation of private health care along socio-economic lines and uneven spatial distribution making it difficult to even predict disease rates due to different risk factors and differing levels of access. This trend can be a bottleneck for any surveillance system. One recommendation is to require private medical providers to report disease cases under the IDSP. Such a mandate should come from the medical licensing authority at the state or local level, and be enforced by the AMC. Ideally, it would be implemented with minimal investment requirements for the providers and be streamlined to encourage participation. In practice, this may prove difficult because of the sheer scale of private care and costs involved. In fact, this is likely a reason it was not initially considered in the IDSP.

Finally, complete health reporting is complicated by reliance on the informal and faith-based sectors for health care. This is not uncommon in impoverished areas and deters any comprehensive surveillance of disease (Duggal & Gangolli, 2005). It is common for patients, who otherwise might be diagnosed at a medical facility, to seek sole medical attention from these providers due to close proximity within slum areas, as well as the lower cost for the visit and medication. This practice foregoes important health data from being included in the IDSP under any capacity (probable, presumptive, or lab confirmed). Most of the time, no specific disease is ever mentioned to the patient due to inability for testing. Patients are commonly released with a pill to treat their symptoms only. These types of visits miss the opportunity to provide valuable health education to a patient and do not provide a safe place for recovery as a proper hospital would. With little oversight, yet high attendance rates, these practices are a detriment to the greater public health. Although a single patient may eventually recover given time, by not being hospitalized or provided with prescription medicine, patients may unnecessarily be exposing a highly communicable disease to their community and households. Increased health education to communities who rely on such practices may help reduce this dependence.

4) Strengthening of spatial methods for outbreak detection

Spatial analyses are often used to carry out outbreak detection (Chen et al., 2011). Due to the spatial nature of waterborne diseases an analysis should be able to identify the location of an outbreak with a reasonable level of accuracy. Currently, the IDSP only allows for analysis at the ward level. As previously stated, this is problematic due to the ambiguous size, shape, and boundaries of city wards and inability to precisely determine the location of a potential contamination source. Formalization of addresses is a very long-term goal. In the near-term, case reports might include the GPS coordinates of the address of each patient. A possible method for this is outlined in the following chapter. The use of either a scan-statistic method, or aggregation of cases into more appropriate spatial units for hot spot analysis to identify clusters also should be considered. Free software packages such as Crimestat and SaTScan can be incorporated to locate hot spots on point data.

In certain cases, it may be more appropriate to use aggregated spatial data at an appropriate spatial unit. For this, a grid analysis with kernel size based on the average size of a city-block is appropriate. This kernel size is selected according to the micro-level scale of risk exposure at the neighbourhood level (Butler, 2006). Such an analysis will provide more appropriate and actionable results identifying hot spots that can be specifically targeted in an efficient manner. Hot spots identified at the city-block sized kernel have a finer spatial resolution. A limitation of this approach, however, is that it only accounts for clusters identified at a hyper-local level. If an outbreak is caused by a more general risk factor, such as contaminated food from a market or festival, this level of spatial analysis may not identify those locations.

3.7 Conclusion and moving forward

In this chapter, I have attempted to answer the following research question: What are the strengths and challenges of the disease surveillance system in Ahmedabad, India, for the control of waterborne disease? Frequent and disruptive outbreaks of waterborne disease warrant this investigation. The current surveillance system, the IDSP, is tasked with facilitating timely and effective public health action for use in health planning, management and evaluating control strategies. A major cause of waterborne disease in Ahmedabad is attributed to illegal water connections and botched repair jobs. This type of activity makes it difficult to consistently monitor high-risk areas for outbreaks. I use the CDC's guidelines for system evaluation to assess the IDSP's ability to deal with this type of public health risk.

From this comprehensive evaluation, I conclude that there is room for improvement in two key areas of surveillance, data quality and timeliness. I identify these attributes as limiting factors inhibiting the ability for disease outbreak detection through geo-surveillance of cases. It is imperative that consistent and accurate case reports for waterborne diseases are aggregated by officials to implement an appropriate response when and where they are required. Presently, there are systematic delays in case reporting, often taking more than three days from an initial diagnosis, which restrict the ability for epidemiologists to carry out up-to-date spatial analyses and unnecessarily delays critical response mechanisms. This is compounded by abundant errors in the report data itself. Up to 44 percent of all case reports are compromised. These limitations within the system are likely to have direct health consequences, mainly affecting the urban poor. Operational changes within the IDSP's operational framework in Ahmedabad should be considered to help reduce urban health inequities by achieving better disease monitoring.

In the following chapter, with regards to the results of this evaluation, I will pilot a prototype method for geo-spatial surveillance using mHealth. In a fully functioning system, robust mechanisms exist for data collection, data analysis, and making those data available to decision makers in an efficient, accurate, and cost-effective process. In the literature, it is shown that a tight integration of information communication technologies (ICTs) for disease surveillance can help achieve these objectives (Cinnamon & Schuurman, 2010; Déglise et al., 2012; Kant & Krishnan, 2010). It is unclear how leveraging spatial methods can help facilitate this. I test how a mobile phone-based data collection tool can help address the issue of data quality and timeliness. The flexibility and scalability afforded by mHealth can be leveraged to improve data collection through forcing standardized data values, allowing data submission directly from the field or health facility, and by being integrated into digital databases for analysis. Due to the ubiquity of mobile phones in India, such technology is available with little cost (DeSouza et al., 2014). Therefore, Chapter 4 will test these recommendations for improving the monitoring of viral hepatitis in Ahmedabad using an mHealth approach.

Chapter 4. Examining the role of mHealth in disease surveillance

4.1 Role of Manuscript in the Thesis

The role of this chapter in the thesis is to describe a case study that examines the role of mobile phone-based disease surveillance. The results of Chapter 3 indicated that geo-spatial disease surveillance in Ahmedabad is challenged by low data quality and frequent delays in case reporting. I develop and test a prototype for mobile phone-based data collection, digital storage of data, and geovisualization of hot spots for viral hepatitis in Ahmedabad. Broadly, the intended role of the prototype is to fulfil and build upon one function of geo-spatial disease surveillance, triggering disease-control interventions (Hashimoto et al., 2000; Thacker et al., 1983) through automated hot spot detection. Specifically, this chapter answers the research question in Objective 2 of this thesis: Can spatial mHealth technology be leveraged for improving waterborne disease control in a low-resource setting through automatic spatio-temporal hot spot detection? Specifically, is it technically feasible for an mHealth approach to be used to initiate disease-control responses through automated hot spot detection to improve disease surveillance?

First, I outline the purpose for mHealth as it relates to disease surveillance and outline the specific objectives the prototype seeks to achieve. Second, I discuss the methodology used for designing and piloting the prototype. Third, I describe results from field-testing of the system. This section is broken into sub-sections by relevant core functions of the system. Finally, I discuss the applicability of the prototype for disease surveillance and automated spatial analysis in Ahmedabad. This research serves as an initial phase of an operational evaluation to test the role of mHealth in disease geo-surveillance.

4.2 Abstract

mHealth is one approach to improving geo-spatial disease surveillance that has been proposed to improve public health activities throughout the developing world. There is little evidence, however, to validate the use of mHealth due to a lack of case studies and metrics used to evaluate its use. I developed a prototype mobile phone-based disease reporting system for identifying hot spots of disease, using viral hepatitis as a pilot case-study. The results from field testing of this prototype indicate mobile phones can be used to reduce the delay in reporting of disease cases. A web-based geographic information system (GIS) was suitable for analyzing, visualizing, and identifying hot spots of disease cases. Implementation challenges that may prevent the use of such a system include training, technical proficiency, cost, and user engagement.

4.3 Introduction

Disease surveillance can be used to identify a previously unknown disease, a risk factor, or the local existence of known risk factors. Spatial disease surveillance is a sub-focus of disease surveillance and used to detect emerging geographical clusters of disease caused by the random occurrence of risk factors in space (Kulldorff, 2001). Spatial surveillance can help identify the local existence of known risk factors by identifying the location of waterborne disease cases which are likely to cluster around a specific contamination source (Chen et al., 2011; Jones & Kulldorff, 2012). The identification of risk factors can then be used to trigger public health response mechanisms (Kamanga et al., 2010; Langmuir, 1963). In urban cities in developing countries, unregulated connections and repairs of the water infrastructure are main sources of contamination for viral hepatitis. Because of the informal nature of this work, it is difficult to monitor the location of such connections or work sites, and thus, the location of this type of risk factor.

The use of real-time data updates and integration of location based services (LBS), and improving cost-effectiveness has been recommended for improving monitoring capacity for public health priorities (Butler, 2006; Labrique et al., 2013; Suresh, 2008). Cinnamon and Schuurman (2010) argue for the application of broader geo-spatial web technologies in disease surveillance for low-resource areas due to their relative cost and ease of deployment. Spatio-temporal analysis methods are particularly well suited for outbreak detection.

Recently, the role of mHealth in public health and disease surveillance has been advocated for by scholars because of advances in architecture and software platforms of mobile phones, built-in GPS capability, and communicative features of mobile phones, in addition to Internet access, which can provide cheap and effective health solutions (P. N. Mechael, 2009; Mehl et al., 2014; Yadav et al., 2010). With the expansion of mobile phone coverage and mobile phone adoption in developing countries, the opportunity for using phones as a health tool (mHealth) has been made possible (S. Patnaik et al., 2009). These interventions rely on the widespread availability of phones, data storage and transmission capabilities including location information[3], and ability for providing quick and cheap forms of direct communication (Labrique et al., 2013). mHealth initiatives currently being developed and tested make use of short message service (SMS), calling, and smartphone applications to record and transmit data for health purposes. Such information, when collected at a centralized location (e.g., stored in a digital database), can potentially be used for geovisualization and data analysis in an automated system. Such a system also could fulfil common disease surveillance objectives, including detecting spatial and temporal clusters (Elliot et al., 2000) and be used as an instrument by decision makers for targeted intervention.

[3] Location transmission can involve triangulation using a GPS inside a smartphone.

To date, case studies on the use of mHealth with a spatial component for disease surveillance are sparse (Labrique et al., 2013; Mehl et al., 2014). The dearth in the literature makes it difficult to examine the role mHealth can play in disease surveillance in developing areas without a typology for evaluation. P. N. Mechael (2009) recommends continued evaluations of case studies to help build up this literature. This paper responds to this call by presenting a case study on the use of mHealth for automated spatial analysis for disease surveillance purposes, using the case of viral hepatitis in Ahmedabad, India. To our knowledge, there are no case-studies of this kind currently published in the peer-reviewed literature.

The objective of this manuscript is to evaluate the technical feasibility of using an mHealth approach to spatial disease surveillance, specifically for hot spot detection of outbreaks. This case study uses a prototype system with mobile phone-based case reporting to achieve automated hot spot detection of hepatitis A and E outbreaks. The prototype was designed to address existing operational challenges to carrying out spatial disease surveillance through the IDSP in Ahmedabad, India. This case study will help answer the research question: can spatial mHealth tools be leveraged for overcoming operational issues in waterborne disease control of low-resource areas?

4.4 Background and context

Study location

Ahmedabad is the largest city in the Indian state of Gujurat with a population of 6.3 million, including 2.5 million whom live in slum or slum-like areas (38% of total population). The city faces a noticeable health divide along socio-economic status, reflective of the national trend (Somani, 2012). Ahmedabad is an appropriate case study for urban health inequity and waterborne disease burden due to the current health profile of the city. Diseases such as acute gastroenteritis, typhoid, viral hepatitis, and cholera are endemic to Ahmedabad, and disproportionately affect the poor wo live in areas where water resource infrastructure is limited (Agarwal, 2011). Slum residents are especially vulnerable to disease through severe overcrowding, deterioration, and unsanitary living conditions (UN-Habitat, 2004). Rapid urban growth in Ahmedabad has resulted in the Ahmedabad Municipal Corporation (AMC) being unable to provide proper water infrastructure, resulting in informal construction and repairs (Saravanan, 2013). The result is open-sewage and widespread illegal water connections, often of poor quality, which cause the mixing of sewage and drinking water. These conditions are considered to be the primary cause of frequent outbreaks of viral hepatitis (hepatitis A and E) centered in slum areas (Arankalle, 2012; Saravanan, 2013). Flood events or contaminated food served at restaurants and festivals are also an important source of viral hepatitis (Arankalle, 2012; Heathcote et al., 2004; Tandon et al., 1985). The high burden of disease, coupled with dynamic risk factors in Ahmedabad make it difficult to detect, diagnose, and control outbreaks of disease until the magnitude is quite large.

Viral hepatitis in Ahmedabad

Viral hepatitis diseases are pervasive waterborne, communicable, and infectious, disproportionately affecting the poor (Purcell & Emerson, 2008) and are endemic to the city of Ahmedabad. Viral hepatitis is spread through ingestion of contaminated water or food through the fecal-oral route, mostly attributed to poor hygiene and sanitation measures. Currently, there is no medical treatment available. Immunizations are available and effective, but, due to cost and access, often the most vulnerable populations go un-treated (Purcell & Emerson, 2008; Tandon et al., 1985). South Asia, and India in particular, bear the burden of these diseases: up to 60 percent of cases and 65 percent of fatalities occur in India.

In this context, in 2008, 233 cases of hepatitis E infection were identified in Ahmedabad, with a case rate of 10.9/1,000 population (Chauhan et al., 2010). Environmental investigation suggested sewage contamination of municipal water in specific locations was the cause for the particular outbreak. This was directly traced to the use of illegal water pipe connections not properly monitored and maintained by the Ahmedabad Municipal Corporation (AMC) (Saravanan et al., 2015). This is a reoccurring source for outbreaks of viral hepatitis in Ahmedabad making it difficult to know the location of high-risk areas for disease in advance. For this reason, I use viral hepatitis as the focus disease of this case study.

The current Integrated Disease Surveillance Program (IDSP) in Ahmedabad

Previous research (Chapter 3) reviewed and evaluated strengths and weaknesses of the existing disease surveillance system in the city of Ahmedabad. Key limitations relate to data quality and timeliness. Instances of multiple entries, missing fields, including addresses or disease information, and records with incorrect addresses, often replacing the patient's home address with the hospital or lab address, render much of the data inaccurate and unusable for analysis. There are also reports of link workers, the community health workers responsible for case reports, not knowing how to submit reports to the central database and regularly missing the task altogether (D. Saxena, personal communication, June 7, 2014). The existing local surveillance system does not have a standard procedure for analyzing spatial data to identify possible locations of outbreaks in a timely manner (Suresh, 2008). In its current state, the IDSP lacks the proper protocols and structure to comprehensively carryout geo-surveillance of viral hepatitis cases in Ahmedabad given the difficulty in identifying and monitoring high-risk areas.

A framework for mHealth evaluation

One expectation for mHealth in developing areas is that it can help facilitate a cost-effective transition to electronic and digital data transmission and storage (Li et al., 2010). Collins et al. (2007) describe the increasing use of EMR, digital databases, and computer based health systems as being beneficial in broader health informatics systems for developed countries. Cost and technical capacity still remain significant barriers to adoption of these technologies in developing areas (Agarwal & Lau, 2010; Cinnamon & Schuurman, 2010; Krishna et al., 2009). The use of mHealth, as well as wider availability of free and open source health tools and the geo-spatial web, have been suggested as ways to bridge this gap, especially for disease surveillance (Cinnamon & Schuurman, 2010; Li et al., 2010; McDonald et al., 2003).

Li et al. (2010) present a case study on the use of geo-tagged mobile alerts to help map the distribution of influenza in Kenya. They describe a method for community health workers to alert regional hospitals on influenza cases from rural villages using a mobile phone. The authors do not use any spatial analysis beyond visually representing the distribution of messages indicating the presence of influenza. In a case study on the surveillance of malaria cases using a similar method, Kamanga et al. (2010) highlight the importance of recording very precise spatial information of cases to properly design an appropriate response. The use of precise location data ensures a more efficient use of resources. In a systematic review on mHealth projects in Africa, Aranda-Jan et al. (2014) report that mobile phone-based data collection results in more accurate and timely reports than paper-based methods. The authors also note, however, the logistical complexity in setting up this type of system is a serious challenge to the long-term and widespread adoption of mHealth. This problem is also highlighted by C Robertson et al. (2010), who comment that the time required to garner support for a mobile phone-based reporting system can be lengthy, meaning intended health outcomes may not change in the short-term.

In this paper, I frame my case study as an evaluation of the ability for mHealth to meet the technical functionality for implementing automated geo-surveillance for hot spot detection of waterborne disease. I evaluate the potential for an mHealth solution to achieve the following disease surveillance functions: facilitate useful and accurate data collection, automate spatial analysis to identify hot spots for disease based on the location of case reports, ensure timeliness of analysis and other operational steps, and ensure effective overall disease surveillance system design. Throughout the literature on disease surveillance and mHealth, these attributes are mentioned as being achievable through mobile phone-based systems (Agarwal & Lau, 2010; Cinnamon & Schuurman, 2010; DeSouza et al., 2014; P. Mechael et al., 2014; S. Patnaik et al., 2009). The specific objectives the prototype aims to be able to achieve include:

1. Decreasing the delay of case reports being submitted and aggregated for analysis. This objective is aimed at improving timeliness in disease surveillance.
2. Ensuring data security and confidentiality in a digital database. This objective is aimed at ensuring acceptability and stability of the system.
3. Collecting accurate and standardized case reports, which include GPS coordinates. This objective is aimed at improving data accuracy.
4. Keeping the cost of the system to a minimum by using free and open source technology. This objective is aimed at improving simplicity, acceptability, and stability of the system.

5. Automated spatial analysis, and subsequent visualization of results, of the distribution of disease to identify hot spots to be used to direct a public health response. This objective is aimed at improving sensitivity and simplicity.

In the following section of this chapter, I describe the methodology used to design and test the prototype. This case study is intended to simulate an initial deployment of the system in an operational situation for monitoring viral hepatitis in Ahmedabad. I intend that such a system would be used by public health officials at the AMC to view, in near real-time, the distribution of cases and possible hot spots of disease using data collected by link workers with mobile phones. This prototype is not intended to replace the IDSP, but rather serve as a case study to provide evidence to evaluate the role of mHealth in disease surveillance by presenting one such approach.

4.5 Methodology

I developed a prototype platform for disease surveillance that responds to existing local strengths, minimizes financial costs and labour intensive procedures, and ensures long-term, accurate and precise data collection to account for weaknesses in the IDSP framework for Ahmedabad. To achieve the design criteria and meet the objectives for the prototype, I used open source and free technology where available such that it does not compromise the core functionality of the system. I also attempted to automate as many of the steps as possible through programming or using existing software packages. The design of the system was informed through consultations with public health practitioners at the AMC, Dr. Deepak Saxena, Dr. Vinayak S. Naik, and Dr. Raja Sengupta, and through evaluation of technical components in the literature (Brunette et al., 2013; Labrique et al., 2013; Slocum, 1999). Decisions about the use of specific technologies to fulfil these requirements are discussed.

Three core-components make-up the framework for the prototype: data collection, data processing and analysis, and data presentation (Figure 5). Appendix B provides a detailed technical description of the prototype's system framework. Data collection addresses the following objectives: standardizing the case report data format, and collecting accurate case reports, which include GPS coordinates. Data processing and analysis addresses the overall design of the prototype and addresses the following specific objectives: keeping the cost of the system to a minimum by using free and open source technology, ensuring data security and confidentiality in a digital database, and decreasing the delay of case reports being submitted and aggregated for analysis. It also addresses the ability to perform automated hot spot analysis on the data. Data presentation addresses the following objective: visualization of the results of hot spot analysis intended for directing a public health response.

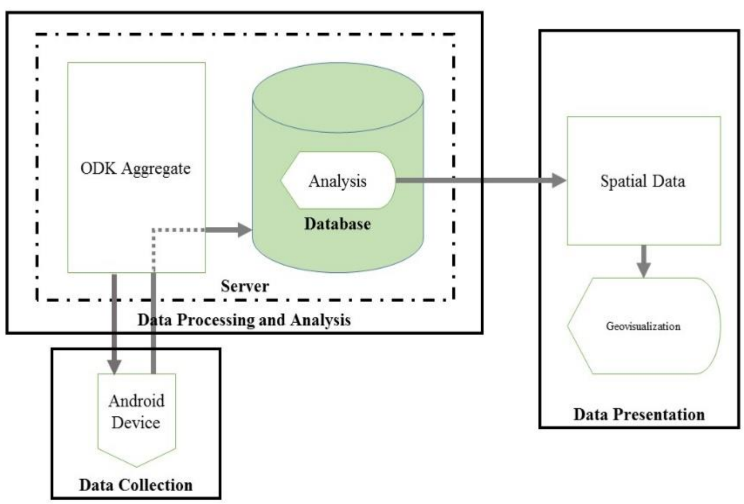

Figure 5. System framework for new surveillance method

In their implementation, each core component is designed to be automated and is programmatically linked to make a coherent and usable system. The only hardware required to deploy the system is a workstation or server with a reliable Internet network connection during data collection hours, and at least one Android device for data collection. Whereas it is preferable that the Android device has a mobile Internet connection (e.g., 2G or 3G) for instant data transmission; this is not compulsory as the data upload can be done at a later time when the device does have connectivity. Confidentiality of the data are ensured through the use of secure logins. Except for home location coordinates, case reports exclude personal identifying information. This methodology was piloted during an outbreak of viral hepatitis in Ahmedabad during November and December 2014. This allowed me to undertake a first phase of evaluation of the system, and compare the data collected through this prototype vis-à-vis data captured within the existing IDSP system.

4.4.1 Data Collection

Data collection was designed to ensure reliable data quality and timeliness. This component comprised of field workers using mobile phone based data collection tools to simulate filling out and submitting case reports. The case reports used electronic forms designed to reflect the same data collected in IDSP reports. Technically, this system utilizes the open source Open Data Kit (ODK) toolbox. The purpose of ODK is to provide a customizable data submission and aggregation tool. I chose ODK because it is free to use and allows for GPS coordinates to be collected. It is appropriate to be used in a low resource environment. ODK also enables Android device-based data collection (Brunette et al., 2013). The open source framework for ODK allows for user-defined data forms or case reports to be administered, and stored, and managed in an electronic database (Brunette et al., 2013). Data collection works by creating an instance of ODK Aggregate, the administration interface for authoring, fielding, and managing mobile data collection, and connecting it to Android devices with the free ODK Collect application installed.

Android is currently the only supported mobile operating system (OS) for ODK collect, the mobile interface used to complete case reports.

Aggregate is hosted on a server and serves as the central administration tool with a graphical user interface (GUI). This GUI allows the system administrator to manage forms, view and track data submissions, and assign security privileges to authorized users. These security features include restricting form access and data submission privileges to data collectors, and limiting access to the collected data. The system administrator would ideally be a public health official serving in a technical capacity. Data collectors then retrieve the data form (a blank case report) to be used onto their Android devices, complete individual instances of the form while in the field, and submit the finalized data to the centralized database. Forms are written in extensible markup language (XML), a markup language that defines rules for encoding data within documents, that can be created manually or through the external ODK Build GUI. ODK Build allows users to graphically create a survey or form using pre-defined question types and implement custom question routing. The questions and responses are customizable with the system administrator defining data response type (e.g., integer and text) and logic restraints (e.g., acceptable value ranges or question routing), as well as allowing for GPS coordinates to be saved within the form. ODK allows for two options to initialize and host ODK Aggregate, either through Google App Engine, or on an Apache Tomcat server with a Structured Query Language (SQL) database. Apache Tomcat is an open-source web server. An SQL database is a relational database that uses SQL for managing and analyzing data. I discuss how both options were implemented and tested below.

a) **Google App Engine option for ODK**

This method of implementation uses Google infrastructure to build and run ODK Aggregate through Google App Engine's Platform as a Service (PaaS). This method requires a Google App Engine account and the Aggregate application is hosted directly on a Google server with a secure URL, available from any Internet enabled connection when provided with the user defined login credentials. The system administrator can assign user privileges for data collection and site administration, and initiate forms for collection.

When data collection begins, and new cases are being uploaded into the database, the system administrator can publish the submissions directly into a Google spreadsheet. A Google spreadsheet is a cloud-based tabular data form, which is part of the Google suite of applications. Some connections require software-to-software authentication. Authentication is required because data are shared between different software components. The objective here is to optimize the data flow, and ensure system simplicity, to an application that allows for analysis. Also, the end-user does not have to worry about manually moving the data from one application to another.

Hosting ODK Aggregate on Google's server platform means data operations are subject to restrictions and daily quotas for the "free" version of Google App Engine. This is one of the challenges in working with a proprietary software platform, particularly one that works with Google's business model that provides managed infrastructure and runtime environments as long as the application follows Google's scope and scale conditions. To customize an application would require additional costs. The application will become temporarily unavailable if any of the quotas are exceeded within a 24 hour period until the next cycle begins. Operations that are limited by quotas include' all requests to read, write, update, or access data in some capacity related to the project. This includes, hosting new submissions, uploading a new form, publishing data to a Google spreadsheet, or downloading the data. This drawback can be avoided by either paying for a higher quota, or by using the Tomcat method of hosting ODK Aggregate. For an operational surveillance system, this method would require a paid service subscription to Google App Engine. There is also a risk of deprecation of Google technologies which would compromise the operation of any system that uses this approach.

b) **Server-based option for ODK**

The second method of deployment for ODK relies on an open source server framework with an SQL-based database. The open source software consists of an information distribution protocol for reading, writing, and hosting data. First an Apache Tomcat 6 server is installed and port forwarding is used to make it visible on a public URL. Port forwarding redirects a communication request from an approved client (data collector) to the host (database) without making the database available to the public. The server can be set up either locally or using free Amazon EC2 cloud services. By using port forwarding, data collectors can submit cases without being connected to the same network as the server, e.g., from a 2G/3G mobile Internet data network. Next, a PostgreSQL database is initialized and configured with the Aggregate installer to host the data submissions. PostgreSQL is an open source and SQL-based relational database management system which easily integrates spatial data formats. The system administrator can access the interface the same way as the Google App Engine method to begin data collection.

Besides the back end of ODK Aggregate, and publishing to Google Drive capability, both methods of deployment allow for the same type of data collection through ODK Collect. This makes the framework transparent to the end user of the application as well as being scalable without changing the front end. The latest version of the application, v1.4.4 is downloadable directly from the Play Store, free of cost. Once installed, the application needs to be configured to connect to the specific ODK Aggregate account for the project to gain access to the form (case report template) using secure credentials. Once a form is finalized, the data are sent into the database directly from the phone and deleted from the phone's internal memory. This step takes less than two seconds.

The purpose of the form, for this case study, is to collect data that would be required to submit a case report in an operational surveillance system. The form was created to test how the prototype can guarantee a standardized reporting structure and capture GPS coordinates. The specific form for this project was designed in collaboration with Dr. Deepak Saxena at the Indian Institute of Public Health. The final set of questions characterize, spatially and temporally, each individual being surveyed with regards to their disease status, while ensuring personal anonymity. In addition to an indicator for type of hepatitis diagnosis, specific dates for the onset of symptoms, official diagnosis from a doctor, and hospitalization were recorded to evaluate how the prototype visualizes the temporal distribution of an outbreak. Auxiliary information such as type of diet and primary drinking water source was collected for descriptive filtering within the geovisualization interface. The exact form that was used is shown in Appendix B.

We tested this system in Ahmedabad for viral hepatitis cases between November and December 2014. This was a suitable case study because of an ongoing outbreak. To test the system, the initial source of viral hepatitis cases came from a spreadsheet of case reports collected in Ahmedabad through the regular IDSP reporting mechanisms during the outbreak. Snowball sampling in the proximate area of each case was used to find secondary cases that had yet to be reported to the IDSP. When the municipal government provided an updated version of the IDSP, we attempted to track down each case using the address and contact information provided in the report. Once the residence of the individual was identified, the form was administered using ODK Collect. In a normal situation, the link workers would learn of new disease cases through their health centre affiliations and locate the residence based on local knowledge of the community (as is currently the case).

Questions from the form were verbally asked, noting down the patient's responses in the proper fields using the touch-screen of the Android device. To collect the GPS coordinates in the form, the internal GPS or location services of the device were turned on. ODK Collect then accessed this data internally and inserted the coordinates into the proper field of the form. Once the form was complete, and data verified, the form was submitted directly from the device while at the residence of the interviewee. As long as there is mobile network data service, the results of the form can be automatically updated into the digital database. If there is no service, the form is uploaded once service is regained, or can be manually uploaded by connecting the device to a computer.

After the form is complete, the fieldworker will ask if any other household members have been diagnosed with viral hepatitis or have experienced any of the major symptoms within the past month. A new instance of the form is used. Next, the fieldworker will ask all household members from residences within a ten dwelling radius of the confirmed case the same question. This decision was made to "search" for new cases. If any individuals indicate a diagnosis or symptoms of viral hepatitis, the fieldworker will ask to see a diagnosis card for confirmation (Figure 6). The diagnosis

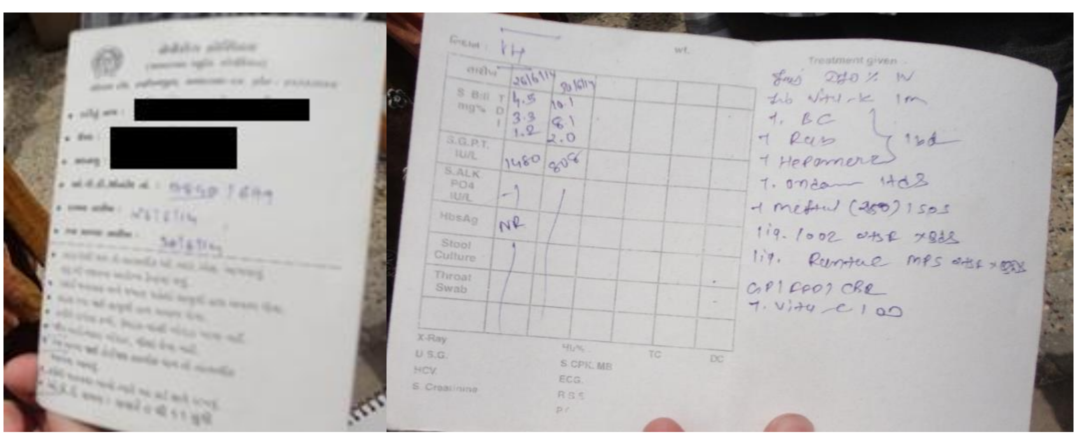

Figure 6. Example diagnosis card confirming disease type

card must have been issued by a medical laboratory or doctor clearly indicating a diagnosis for the case to be reported in the prototype.

4.4.2 Data processing and analysis

Once data collection is ongoing, the results are continuously streamed into the centralized database. From here a series of trigger functions and python scripts were written to: 1) create a series of KML files based on specific outcomes and attributes, and 2) create a series of KML files showing the results of hot spot analysis. This procedure was designed to ensure analysis is accomplished swiftly and consistently. Python coding is ideal for this task because of its strong support for integration with other technologies. The procedures for each of these two tasks are presented below.

a) Creating individual case spatial layer

First, individual case data from the database are converted into informative and descriptive KML files to be displayed as layers in the final geovisualization. As data are streamed into the database, a trigger function listens for new data updates. If an update is detected, a function is executed to create a new point layer representing disease cases for the geovisualization. For the sake of privacy and confidentiality of the individuals, these KML files remain on the local system.

b) Automating spatial analysis

The second step in the analysis consists of using GI* to determine if any hot spots exist across the city of Ahmedabad based on the number of cases occurring within each city-block sized kernel. GI* was used to provide direct comparison with the potential spatial analysis of IDSP data. This analysis is done using the PostGIS extension in PostgreSQL, and GeoDa, the python extension for GeoDa. PostGIS and GeoDa were used because they can be implemented programmatically, are free and open source, and output cross-compatible spatial data file formats. Each time the geovisualization interface is opened the spatial layers present the most recent overview of the distribution of cases along with statistical significant GI* values for each ward. In an operational system, the specific algorithms used for spatio-temporal analysis can be tailored to the specific context.

All cases diagnosed within a one-month period are considered for analysis. The hot spot analysis will identify spatial units with a statistically significant high rate of incidence compared to what is expected, and the usual levels throughout the rest of the city. For comparison, all kernels are included in the KML file with the number of cases, as well as general demographic information written as text attributes for each KML feature. The date of the spatial analysis is also included as a timespan element in the KML.

4.4.3 Data presentation and geovisualization

The final process in the system is the geovisualization interface intended to be used by public health practitioners. This tool shows the location of each case, and any areas that are statistical hot spots (Figure 7). This information can be explored temporally and spatially in an interactive method to compare attribute information for each feature. It is an HTML-based webpage that uses the Google Maps API to overlay the KML layers, and uses JavaScript coding to allow interactive functionality and temporal filtering.

I decided to include a temporal visualization of the data because the temporal distribution of disease can be important for disease surveillance (Mantel, 1967). In an operational surveillance system, temporal cluster methods should be used. I did not include statistical temporal analysis in the prototype, however, the user can filter the map by time period to make visual conclusions on the temporal trend. I chose the TimeMap JavaScript package to create a time slider and enable temporal filtering of the data within the map. I chose this package over other packages because it automatically detects temporal information which is not a native function of the Google Maps API. This package accesses the timestamp elements of the features within the KML and only shows the features with a time value within the specified temporal period. I set the temporal scale to a week but the period can be adjusted by the user.

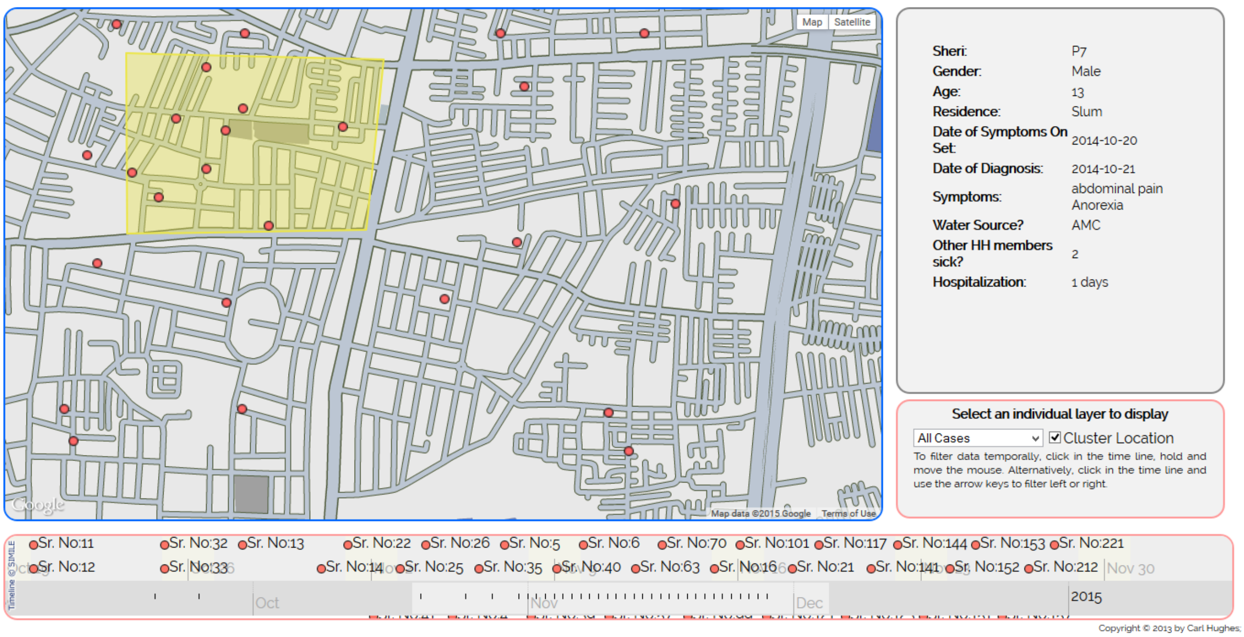

Figure 7. Geovisualization interface with anonymized individual cases and hot spot location identified within yellow polygon

I chose to activate the KML files via a drop-down menu. Each menu option links to a unique KML file based on the specific hepatitis status or other attribute including presence of symptoms, age, gender or if other household members have hepatitis-like symptoms. The user can click on feature icons from within the map or timeline to view their attributes. The goal here is to present information to the user that may be useful for deciding on an appropriate public health response. Since the process for KML creation is automated, as new data are collected and the KMLs updated, these changes are reflected in the geovisualization interface in near real-time.

4.4.4 Field Testing and Evaluation

Field testing (Figure 8.) consisted of three steps related to the core functions of the prototype. First, data collection was tested using field-workers to simulate link workers completing and submitting case reports and is evaluated as the ability to collect standardized and accurate case reports including spatial information. Data collection relates to timeliness and data quality. Second, use of the digital database and automated analysis functionality was used to evaluate the applicable data flow. Data processing and analysis relate to sensitivity, flexibility, stability, and simplicity of the prototype for conducting analysis. Finally, the geovisualization was tested to see how the prototype is able to present important findings from analysis to direct response efforts. Data presentation relates to stability and acceptability. Because this is an initial prototype and not an operational system, applying the same framework used in Chapter 3 informed by CDC guidelines for system

Figure 8. Field testing of the disease case data collection

evaluation is not applicable to this section. Instead, I evaluate the prototype on the ability to fulfil the given system objectives (Table 8).

Table 8. Methodology for prototype evaluation by system objective

Objective	Method for Evaluation
Decreasing the delay of case reports being submitted and aggregated for analysis	I evaluate Objective 1 as the ability to reduce delays in data reporting from within the system. Objective 1 is related to timeliness. Effective disease interventions rely on the rapid transfer of information regarding the trends of diseases. Delays in reporting of waterborne viral hepatitis, in the example of Ahmedabad, can affect health outcomes for populations relying on drinking water that is potentially contaminated.
Ensuring data security and confidentiality in a digital database	I evaluate data security and confidentiality as the ability to retain data safely within the system and to prevent its unauthorized access. Objective 2 is related to acceptability and simplicity.
Collecting accurate and standardized case reports, which include GPS coordinates	I evaluate data collection as the ability for the mobile phone-based system to complete and submit instances of accurate case reports with GPS coordinates in a timely manner. Objective 3 is related to data quality and acceptability.
Keeping the cost of the system to a minimum by using free and open source technology	I measure cost as the required investment for all components within the system. Objective 4 is related to acceptability and simplicity. One of the primary challenges this system addresses is the high cost of disease surveillance through ICTs that often poses an insurmountable barrier for disease monitoring in developing countries.
Automated spatial analysis, and subsequent visualization of results, of the distribution of disease	I evaluate analysis and visualization as it pertains to data flow and the ability to each step in the process as easily as possible, especially hot spot analysis. Objective 5 is related to sensitivity, simplicity, and stability.

4.6 Results

The prototype reduced the time between data collection, analysis, and presenting the results to public health professionals (Table 9). This was accomplished through the use of automated methods for data storage, analysis, and visualization. The time required for a new case in the pilot study to be submitted from the field, updated in the database, re-analyzed with previous data, and presented along with hot spot results in the geovisualization interface, was less than one minute. This compares from the existing method which takes, on average, over one day for a case to be submitted into the database. This speed in data flow can be achieved even with the most basic connectivity specifications: a 2G data connection in the field and a 5mbps Internet connection on which the server is being hosted. The speed will vary by computer capacity and number of cases. As the database increases in size, the analysis will take longer. Even then, with linear scaling it is estimated that the process will not exceed a reasonable time frame, for example, five minutes, with 10,000 records.

Table 9. Results from field-testing of prototype system

Objective	Result
Decreasing the delay of case reports being submitted and aggregated for analysis	Submission of case reports to the database was almost instantaneous (several seconds)
Ensuring data security and confidentiality in a digital database	Data confidentiality were ensured for all cases through assigning user credentials
Collecting accurate and standardized case reports, which include GPS coordinates	Every case report collected through the prototype adhered to the defined format and data requirements Accurate GPS coordinates were recorded
Keeping the cost of the system to a minimum by using free and open source technology	Estimated at 5,518 USD for a 12 month deployment
Automated spatial analysis, and subsequent visualization of results, of the distribution of disease	Automated hot spot analysis was achieved and visualized, in near real-time, in an interactive geovisualization

Through the use of security credentials, the data are protected and confidential once submitted from the device. The device itself can be password protected to prevent access to any un-submitted data. The system administrator, ideally a technical public health official from the DSU, has the power to limit data collection to authorized users. The system administrator can see in the database the contributors of each entry to verify if the proper data collectors are submitting the correct reports. Access to ODK Aggregate itself is also password-protected and can be restricted to access from certain IP addresses. The KML files that are created are kept on a secure local password protected computer and are not available online. Access to the geovisualization interface is also password protected to authorized users. As the KML files are located on the one local computer, the geovisualization interface can only be used on that machine. These restrictions prevent the public or malicious parties from accessing confidential residential locations of individuals.

For data collection, a 2G SIM card with 200MB data limit was able to meet the necessary mobile data requirements for an Android device submitting results over the month long trial. I tested the prototype on the AirTel and Reliance telecommunication networks. More specifically, the 2G connection was used for downloading ODK collect on the device, connecting to the central ODK Aggregate account and downloading the blank case report template, and submitting data into the database. In total, less than 50 megabytes of data was used for these steps per device to record approximately 35 submissions each.

Critically, the 2G network coverage was constant and reliable for the entire duration of field use. By constant and reliable, I mean that uploads and downloads never dropped and data submission was possible at all times during the field-testing phase. The time it took for one completed form to be added into the database was 5.6 seconds. Consequently data submission was possible from each household location that was visited. Finally, despite a significant source of potential signal interference in dense urban settings, the performance of the GPS on the Android device was good with an average accuracy of 8.1 meters and taking less than five seconds to achieve this precision while completing the form. Each case report was completed according to the defined form requirements which meant all cases followed the same standardized data scheme. In practice, it may be hard to verify if link workers are inputting the correct response (even if it is an allowable response) or chose to cancel a report altogether because of the enforced data scheme. The prototype was able to improve the timeliness and data quality of data reporting compared to the IDSP dataset I evaluated for Chapter 3.

The results of this study suggests that the back-end of the system can be implemented with minimal cost by utilizing open source software and existing infrastructure where available, without limiting functionality. Open source software in the prototype were used to create the server framework to host the database, automate spatial analysis, and invoke interactivity in the geovisualization.

Table 10 provides an estimated breakdown of the associated costs for each component in the system, extrapolated for a 12 month deployment.

Table 10. Breakdown of technical costs associated with surveillance system

Item	Purpose	Cost (USD)	Notes
Android device	Data collection device	74/unit	
2G Mobile Service	Near real-time data collection	52	
Server	Host ODK Aggregate	0-195	
Link workers	For Data collection	336/worker	Already employed by AMC

Computer	Administration purposes and for use of geovisualization tool	500/unit	Existing access through AMC
Internet Connection	Set up ODK, run server, and view geovisualization	660	Existing access through AMC
ODK		0	
Database	Host data	0	Free with PostgreSQL option
Python coding	Software for analysis	0	GeoDa, SimpleKML, Geovisualization
System administrator	To maintain technical components of system	5,000	

Many of the associated costs for deploying this system may be absorbed in the existing operating budget for the AMC. The link workers who would be used for data collection with the Android devices are already responsible for writing case reports to the IDSP, among other general health activities as sanctioned in their respective community boundaries. Their wages are currently being paid for by AMC. The cost of training was not measured in this case study and is likely to be a reoccurring expense in an operational situation. A system administrator would also need to be hired to set up and maintain the system. This likely represents the biggest financial burden for a public health agency adopting the technology.

A computer with an Internet connection can reasonably be assumed to be covered by the AMC as this technology is also already required for IDSP, as well as other day-to-day tasks of the health department. Multiple options exist for hosting ODK aggregate on an online server. These range from a free-of-cost Apache Tomcat 6 open source web server to a cloud-based Amazon EC2 subscription for 195 dollars per year. This study utilized the open source option as all security and data needs were met (Saravanan, 2013) without additional costs.

As such, the only operating expenses for deployment are the costs for the Android device and a data plan, and personnel costs for setting up and maintain the system. Fortunately, the Android costs in the Indian context continue to decrease with fierce market competition and innovation (Doron & Jeffrey, 2013; S. K. Singh, 2008). As ODK can run on any Android device running Android 2.2 or higher, the most basic, and thus, cheapest technology is adequate. Tablets are approximately $74 per device and data plans are as low as $52 per year. These costs can potentially be mitigated by using a cell phone instead of a tablet at a lower cost per device or even making use of link workers' personal devices with a possible subsidy from AMC. Such a solution may create a heterogeneity problem where different link workers have different levels of access and proficiency with the technology. As the devices could be utilized for a minimum of two years, the yearly cost per device can be further reduced. The expected technology costs for utilizing this system with one device for each of the 62 wards (assuming a two year device life) over a 12-month period would be approximately $10,518. This estimate represents a theoretical additional operating cost for integrating the prototype beyond existing IDSP expenses in Ahmedabad. In the future, cost-benefit analysis can be used to predict the actual value-added of this approach given the broader disease burden (e.g., changes in response resources, preventive action, impact of disease outbreaks).

The overall design and data flow in this prototype was able to ensure automation, standardization and confidentiality of the data. By using mandatory fields in the form, all data collectors are required to collect a minimum amount of information for each case required for accurate cluster detection. Using the input restraints, answers to certain questions can also be validated directly as they are filled in. The data collector is also prompted to correct a response if it is outside of the acceptable input range. The result of this functionality means that each entry in the database has the same database structure and they can be used for the analysis without interrupting the automated process. This ability is also apparent in the geovisualization interface where the user can directly compare the attributes for individual cases (or ward-based hot spots) because the data for each case follows the same format.

Hot spot analysis requires timely reporting of precise and accurate location data. In the current disease reporting system, data are updated infrequently and often cases are missed or incorrectly reported. This results in entries missing the patients' home location, or containing incomplete or incorrect home locations. Often the hospital or doctor's address is listed and is mistakenly used in analysis. These erroneous entries are often compounded with other errors such as multiple entries of the same case, possible missing entries, and errors in disease diagnosis or errors in transcription. Low quality directly influences the ability to do analysis for outbreak detection.

Because the prototype captures the GPS coordinates of the data collection device, geocoding of precise and certain addresses can be ensured (i.e., given that the link worker is standing in the correct location). Link workers are embedded in a specific neighbourhood and already conduct regular house visits. Similar mHealth systems with house visits have been tested in urban areas for maternal health purposes using mobile phones for point-of-care diagnostics and decision support systems (S. Kumar et al., 2013; Labrique et al., 2013; Mehl et al., 2014). By utilizing the GPS information, I found the system was able to get the coordinates of the individual's residence without having to write the address. The device required fewer than three seconds to record accurate coordinates with fewer than 20 metres of error, and does not require any technical GPS expertise as I wrote the application to capture and then update the coordinates directly into the correct field in the form.

Collecting GPS coordinates is not without problems. *Chawls*, which are informal settlements, are characterized by narrow alleyways with a high density of medium rise buildings. This configuration can be a potential source of satellite interference. The accuracy of the GPS data appeared not to be affected as evident from similar accuracy rates across the city with different levels of potential satellite interference. In collecting the data directly from the field, the accuracy of the database, and subsequent analysis is improved. By knowing a more accurate location, the number of cases in each spatial unit can better enable statistical hot spot analysis. Having locations collected in-situ allows for more advanced post-hoc clustering analysis of datasets, such as applying the spatial scan statistic or Bayesian mapping methods. Therefore, the prototype demonstrates one method to collect spatial data that can be used in an automated hot spot analysis.

4.7 Discussion

The field-testing of this technology demonstrates one viable method for beginning to account for challenges facing disease surveillance of waterborne disease in Ahmedabad. These challenges include accuracy, timeliness, data analysis, and data management. Together, these challenges affect the timely identification of statistically significant hot spots of disease and consequent interventions from being targeted and implemented properly to reduce risk factors for viral hepatitis. The results also provided evidence that the prototype's key objectives were achieved. These objectives were selected to examine the technical feasibility of using mHealth for geo-spatial disease surveillance through automated hot spot detection of disease outbreaks. This case study also supports claims made by Cinnamon and Schuurman (2010) for the broader use of geo-spatial web technologies for disease surveillance in developing countries. This case-study did not consider the broader role of mHealth, outside of geo-spatial disease surveillance, in public health, for example, triggering disease control interventions or improving health outcomes.

This prototype aims to address challenges facing the IDSP in Ahmedabad through low cost technology, capitalizing on existing local strengths, and integrating emerging technologies that are often only used in isolation for mHealth applications. A core strength of the existing infrastructure is the link workers employed by AMC. They possess extensive local knowledge and are already mandated to report disease cases (albeit under the IDSP framework). As such, reporting with an Android device should not detract from their other responsibilities, such as acting as referral agents, promoting health skills, and providing health education (Barua & Singh, 2003). It is possible that they will have more time for completing other activities as the time spent doing Android-based reporting is significantly less than the existing paper-based reporting. Conversely, increasing productivity may be faced with resentment from link workers. The case study did not evaluate the willingness for link workers to engage in this technology. In fact, the use of mobile phones by link workers in an official capacity may result in some of the same acceptability issues of the IDSP described in Chapter 3, for example, incomplete or incorrect case reports, and not submitting case reports at all. Despite replacing a paper-based method, if link workers remain in dispute with the AMC, many of the same acceptability issues would remain.

Deployment is strengthened by the fact that cell phone technology adoption rates are growing rapidly. Link workers would require training on the system. This may be simplified if they have prior experience using Android devices. Additionally, much of the backend technology required to operate the system can be reasonably expected to be available within the existing IT infrastructure of the AMC. This is further mitigated because the actual technology requirements are relatively low-tech (e.g., one basic computer with the Internet and power connection). By embedding the system in the existing IT infrastructure, the task of setting up and maintaining the server is also simplified.

It is noted, however, that configuring the software dependencies (e.g., digital database, Python/JavaScript coding, and APIs) is more complex than the hardware integration. Due to the nature of free and open source software, the user needs to have some experience in the specific programming language of the software to design the processing and data flow functions. Although the software is quite flexible in its utility, each application requires customized programming and distribution systems that are commonly not provided or pre-packaged for each specific user. Furthermore, a system administrator would need to be available at all stages of data collection to troubleshoot the system or implement changes in the form for case reports. More broadly, implementing a change in the data scheme for the case reports would require retraining data collectors, and recoding the system to account for the different data structure. The time and resources this would take to achieve are unclear as it would depend on the type and scale of the change.

A second major challenge that the system is able to address is its ability to rapidly disseminate critical information necessary for public health responses by decision makers. From Chapter 3, I conclude, under the existing IDSP system, reporting is intermittent and requires several error-prone manual steps to aggregate the data, analyze the data, and then present this data in a coherent way. The problem is further exacerbated by the lack of a clear and structured procedure. Aggregating paper-based reports into one electronic database is complicated from having no standardized data scheme for case reports, frequent delays in reporting, and a rudimentary aggregated database using a Microsoft Excel spreadsheet.

By amalgamating and automating several of these steps, this system is able to directly link the data collection with the presentation of results in near real-time. This prototype also enforces a standardized data scheme by forcing specific data types and inputs from the field and using a cloud-based server to aggregate the data into a structured database. This structure allows data analysis to be done programmatically and automatically as data are updated. The time between data collection and visualization is significantly reduced; data are more reliable; and any technical GIS or statistical analysis is done automatically. Decision makers have direct access to more accurate data in accordance with accepted public health standards and have more time to create and implement interventions based on the ability to identify with credible spatial accuracy, hot spots for disease. These functions indicate strengths in terms of timeliness, stability, and data quality.

Although this system was tested on hepatitis A and E in Ahmedabad, India, it is not limited to any one particular disease or geographic location. The system is designed to be adopted for any type of spatial monitoring where cost and resources are limited and near real-time reporting is essential. This is possible through the flexibility and low cost afforded by use of open source software, the ease of deployment and maintenance of ODK, and ability to programmatically customize components in the system.

In terms of data collection, the specific form to be used is customizable, meaning it can be tailored to the project’s specific needs. By using a PostgreSQL database, advanced spatial analysis can be implemented using PostGIS, OGR, GeoDa or other GIS modules with technical user-defined algorithms for many types of spatio-temporal methods. Therefore, the implementing agency will require technical proficiency at their desired level of analysis. The results of the analysis can also be converted into different spatial formats to be used in various visualization methods through different spatial APIs or programs including ArcAPI, Open Street Maps, and CartoDB. This is in comparison with a basic Microsoft Excel based database which would require several manual steps to achieve the same analysis. PostgreSQL is an ideal database framework because PostGIS is not compatible with other SQL-based databases.

The prototype can be deployed in emergency situations for short and long-term monitoring. It takes less than one day to initially set up the framework, given the existing hardware and technical capacity. The speed for deployment make the system suitable for use in sudden or unexpected situations such as emergency disease outbreaks, conflicts or environmental disasters. Once the system is set up, it is possible to ensure long-term use and maintenance by using a digital database, which can be done remotely. With ODK, the system administrator can allow multiple and simultaneous data collectors maximizing the reach of data collection. A database server instance of ODK can store very large data sets of over 100,000 records. Without paying for the premium service, the Google App engine instance of ODK has a limit of 50,000 read or write operations which is quickly exceeded, even when streaming less than 100 data entries to a spreadsheet during ongoing data collection.

The novelty of automating the analysis process in such a situation is that it bypasses the need for the technical expertise and analysis from a GIS professional, which can be a limiting factor in a resource-scarce setting or emergency situation (A. Singh et al., 2011). The requirements for implementation include a computer with basic market specifications and an Internet connection, and one or more Android devices. By storing the data in a secure database, it is possible to do ad-hoc analysis for further insight. Data security will ensure the data remains in place and is easily accessible after initial analysis.

Throughout the field-testing of the prototype, several operational and theoretical issues became apparent, complicating deployment. The most obvious issue that may limit implementation concerns the technical requirements for initially setting up the system, and subsequent data collection. To create the ODK account and set up the database requires someone with knowledge of database and credential management. This limitation has been reported in the literature as a common drawback for mHealth in developing areas (Estrin & Sim, 2010; P. N. Mechael, 2009). This criticism of mHealth is currently exacerbated in developing areas by the existence of the digital divide.

Related to the previous issue is the selection and implementation of specific algorithms for analysis. Although the GI* is used, this may not be practical or applicable in all situations. In fact, it does not account well for point-based cluster analysis. A second major drawback in this prototype, as it stands, is the inability to account for temporal clustering methods. Currently, pre-packaged or open source methods for point-based and purely temporal or spatio-temporal analysis are not readily accessible for integration with Python and would require complex and highly technical coding. Some of these functions can be achieved through stand-alone or pay-per-use packages, however, this would increase the cost and complexity of the system.

Next, the education and training of link workers can limit data collection. Link workers will be required to adopt a technology with which they may have had no prior exposure. Although the survey interface is designed to ensure usability, certain issues that may compromise the data include: failing to turn on the GPS, selecting an incorrect response, not submitting a completed form, or overall discomfort in using a smartphone application. Adding to this limitation is the rate at which mHealth technology is changing (Tomlinson et al., 2013). Any data collectors will need ongoing formal training sessions to keep up with changing technology. Paper-based reporting does not require this level of engagement and continued investment. Fortunately, the ability to deploy the system in the local language and the rapid diffusion of mobile technology at-large further help address these issues where use of the technology and training could be eased.

Another potential challenge for deployment is ensuring proper use of the mobile phones by health workers. In previous studies, personal use of the mobile phones, including playing games, email, and social media, have depleted resources by using up data allotments (D. Saxena, personal communication, June 15, 2014). This complicates the potential decision that sponsoring agencies would need to make on providing, either, use of "company" phones or subsidies for personal phones in a given program. Remote monitoring of the phone's usage would likely cause privacy and trust issues between data collectors and system administrators. Furthermore, without frequent and comprehensive monitoring of the health workers and follow-ups with diagnosing facilities, it would be difficult to check that each diagnosed patient receives a house visit and has a case report filed. Missing cases would reduce the representativeness of the system and interfere with outbreak detection. This concern applies equally to the paper-based method.

Finally, in its current form, this type of approach cannot easily integrate the private sector. The willingness of private providers to submit case reports according to the same framework was not measured within this case study. Private providers are important stakeholders in disease surveillance their inclusion in public health activities would be beneficial, however, the likely costs involved in participation may deter engagement. This issue also relates to the complexity of public-private partnerships where conflicting interests have so far prevented the scalability of mHealth (Barlow et al., 2013; P. N. Mechael, 2009). Further research, including scaled program evaluations, is required to more conclusively comment on these challenges.

4.8 Conclusion

In this chapter, I present a prototype and initial phase of field-testing for that prototype. This prototype is an example of an mHealth approach to improve near real-time disease surveillance in Ahmedabad. The prototype consists of Android device-based data collection, automated data storage and analysis for hot spot identification, and an interactive geovisualization interface with a data flow for decision makers. This prototype is able to rapidly identify spatial hot spots of disease by identifying significant clusters of existing cases as they are diagnosed and reported. These hot spots can represent areas of populations who are at risk for disease. In theory, public health officials will use the tool to create and deploy targeted interventions for reducing further cases. The tool overcomes barriers to accurate data collection and reporting, technical and labour-intensive analysis, and rapid communication of the location of any hot spots of disease in a near real-time and cost-effective way.

It should be noted, however, that this prototype does not conclusively validate the use of mHealth for disease surveillance. Issues that remain in place and may prevent improvements in health outcomes include: being overly dependent on technology in a resource scarce environment, training and usage issues related to user engagement, the sustainability of the technology, and the inability to carryout operational monitoring of the system. Finally, it is important to mention that this prototype only presents one sub-application of disease surveillance, spatial hot spot analysis of disease cases.

The results of field-testing suggest public health practitioners can benefit, in certain aspects, from mHealth. Importantly, the first function of public health, as defined by WHO (2010): "The assessment and monitoring of the health of communities and populations at risk to identify health problems and priorities" are facilitated through identifying hot spots of disease, and subsequently potential contamination sources, with improved spatial and statistical certainty. The approach I present is novel in the way it integrates the three core components of disease surveillance: data collection, analysis, and communication of results into a structured and automated system capitalizing on mobile and open source technology. This case study represents one of the first times that an open source mHealth tool has been used in this way and for this purpose.

Chapter 5. Conclusions

This thesis evaluated the current disease surveillance system in Ahmedabad and prototyped an mHealth tool for case reporting with spatial analysis and geovisualization for waterborne disease hot spot detection. The thesis directly addresses the issues raised in the literature for leveraging mobile and spatial technologies for better public health monitoring (Chang et al., 2009; S. Kumar et al., 2013; Mehl et al., 2014; Vyas et al., 2010). Two primary questions were answered in this thesis: (1) What are the strengths and challenges of the existing disease surveillance system in Ahmedabad, India, for the control of waterborne disease, and (2) can spatial mHealth tools be leveraged for improving waterborne disease control in a low-resource setting?

The first question was concerned with providing a comprehensive evaluation of the IDSP in Ahmedabad according to a review of system attributes proposed by the CDC (German et al., 2001). The purpose of the evaluation was to assess how the IDSP can identify likely populations and locations exposed to risk factors. The following system attributes were assessed: simplicity, flexibility, data quality, acceptability, sensitivity, predictive positive value, representativeness, timeliness and stability. Evidence was collected through an analysis of four months' worth of anonymous case reports, in addition to a review of progress reports and status updates from funding bodies and stakeholders involved with implementation of IDSP (World Bank, 2012a, 2012b).

The analysis suggests the IDSP can benefit from improvements in the way data collection is carried out. The analysis also highlighted the limited capacity for carrying out spatio-temporal hot spot detection. This function is important given the fact that health risks are likely to be spatially clustered. Specifically, data quality was found to be compromised through systematic errors including missing and/or ambiguous field values, multiple entries of case reports for the same health event, and failure to adhere to standardized reporting formats and values including contact information and ward names. The overall impact of IDSP is further negated by frequent delays in case reporting. Case reports often took over three days from the initial diagnosis. Delays in cases being submitted to the DSU means outbreak detection and response may be critically delayed. It also means otherwise healthy persons may be exposed to preventable risk factors before an outbreak is actually detected. This evaluation reflects concerns shared in the literature on disease surveillance of public health risks from delays in case reporting (Jajosky & Groseclose, 2004; Overhage et al., 2008; Wagner et al., 2001) and data quality issues (Berkelman et al., 1994; Butler, 2006; German et al., 2001).

The existing process for case reporting also impacts analysis. The ability for spatially accurate hot spot analysis to identify outbreaks is problematic given the location information collected in case reports. GIS methods are limited to identifying an outbreak at the ward level. Due to MAUP-related issues and the nature of waterborne diseases, this level of analysis is inadequate for deploying targeted and efficient public health interventions to contain an outbreak.

The second question is concerned with testing the use of mHealth's spatial functionality for public health. This thesis prototyped an mHealth tool for improving specific components of data collection, outbreak detection, and intervention planning, keeping in mind the issues with IDSP presented earlier. The prototype uses open source software for Android tablets and mobile phones, with customized forms to collect case reports, followed by direct, automated submission to a cloud-based database from the field. This prototype uses free and primarily open source tools for the database. A hot spot analysis is then automatically run on this data to provide an updated cluster map as a consequence of ongoing data submissions from the field. The results of the analysis are displayed on a geovisualization interface for public health officials. This system was designed to (1) decrease the delay of case reports being submitted and aggregated for analysis, (2) provide data security and confidentiality in a digital database, (3) collect accurate and standardized case reports(which include GPS coordinates), (4) keep the cost of the system to a minimum by using free and open source technology, and (5) automate spatial analysis and subsequent visualization of results showing the distribution of disease and possible high-risk areas for further disease.

Field-testing indicated both strengths and weaknesses of the technology. The most relevant benefits from this system vis-à-vis the IDSP are (1) the ability to ensure standardized and accurate data collection through implementing value restraints and logical question routing, (2) the rapid reporting time from form completion to the data being included in the database and its subsequent hot-spot analysis, and (3) the seamless integration of collecting precise location coordinates for each reported case from the internal GPS of the device. Overall, the prototype was able to improve data quality and reduce the delay in case reporting, and consequently improved the ability to carry spatial analysis.

Conversely, practical and theoretical challenges to implementing the tool were found. The dependence on technology for initially setting up the system, and its actual operation may complicate scalability. User engagement is weakened because of the need for regular training and lack of monitoring metrics. Additionally, the use of the GI* has very narrow public health use. Integration of more complex spatio-temporal analysis methods may help broaden the public health benefit for Ahmedabad. This integration would however require increased technical capacity.

This thesis contributes towards an emerging literature on mHealth and provides insight into an application with integration of mobile data collection, spatial analysis, and geovisualization. As governments and organizations continue to look towards long-term public health interventions, continued research into the specific functions and benefits of mHealth, and more broadly ICTs, is warranted. As these technologies are considered elsewhere, it will be important to reflect on local contexts including the availability of technology, nature of the public health objectives at-hand, limiting factors such as the digital divide and training, and existing operational challenges in need of being addressed. Fortunately, the relative ease with which these technologies are leveraged is increasing through the use of open source code, widespread availability of ICTs, decreasing operating costs, and cross-platform and device compatibility.

References

Agarwal, S. (2011). The state of urban health in India; comparing the poorest quartile to the rest of the urban population in selected states and cities. *Environment and Urbanization, 23*(1), 13-28.

Agarwal, S., & Lau, C. T. (2010). Remote health monitoring using mobile phones and Web services. *Telemedicine and e-Health, 16*(5), 603-607.

Ahmedabad Municipal Corporation. (2008). *Statistical Outline for 2006-07*. Ahmedabad Retrieved from http://www.egovamc.com/corporation/frmstatisticaloutline.aspx?sno=StOutline&yr=2006.

Alwan, A., MacLean, D. R., Riley, L. M., d'Espaignet, E. T., Mathers, C. D., Stevens, G. A., & Bettcher, D. (2010). Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. *The Lancet*, *376*(9755), 1861-1868.

Aranda-Jan, C. B., Mohutsiwa-Dibe, N., & Loukanova, S. (2014). Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. *BMC Public Health*, *14*(1), 188.

Arankalle, V. A. (2012). Hepatitis E in India. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences, 82*(1), 43-53.

Arts, D. G. T., de Keizer, N. F., & Scheffer, G. J. (2002). Defining and improving data quality in medical registries: A literature review, case study, and generic framework. *Journal of the American Medical Informatics Association*, *9*(6), 600-611.

Balarajan, Y., Selvaraj, S., & Subramanian, S. (2011). Health care and equity in India. *The Lancet, 377*(9764), 505-515.

Barlow, J., Roehrich, J., & Wright, S. (2013). Europe sees mixed results from public-private partnerships for building and managing health care facilities and services. *Health Affairs, 32*(1), 146-154.

Barua, N., & Singh, S. (2003). Representation for the marginalized—Linking the poor and the health care system: Lessons from case studies in urban India. New Delhi: World Bank.

Beratarrechea, A., Lee, A. G., Willner, J. M., Jahangir, E., Ciapponi, A., & Rubinstein, A. (2014). The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. *Telemedicine and e-Health, 20*(1), 75-82.

Berkelman, R. L., Bryan, R. T., Osterholm, M. T., LeDuc, J. W., & Hughes, J. M. (1994). Infectious disease surveillance: a crumbling foundation. *Science-AAAS-Weekly Paper Edition, 264*(5157), 368-370.

Binnendijk, E., Koren, R., & Dror, D. M. (2012). Can the rural poor in India afford to treat non-communicable diseases. *Tropical Medicine and International Health, 17*(11), 1376-1385.

Breslauer, D. N., Maamari, R. N., Switz, N. A., Lam, W. A., & Fletcher, D. A. (2009). Mobile phone based clinical microscopy for global health applications. *PloS One, 4*(7), e6320.

Brunette, W., Sundt, M., Dell, N., Chaudhri, R., Breit, N., & Borriello, G. (2013). *Open data kit 2.0: expanding and refining information services for developing regions.* Paper presented at the Proceedings of the 14th Workshop on Mobile Computing Systems and Applications, Jekyll Island, Georgia.

Butler, D. (2006). Disease surveillance needs a revolution. *Nature, 440*(7080), 6-7.

California. (2011). *Title 17, California Code of Regulations (CCR) §2500, §2593, §2641.5-2643.20, and §2800-2812 Reportable Diseases and Conditions*. (110a). Retrieved from http://www.cdph.ca.gov/HealthInfo/Documents/Reportable_Diseases_Conditions.pdf.

Carneiro, H. A., & Mylonakis, E. (2009). Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases, 49*(10), 1557-1564.

Chang, A. Y., Parrales, M. E., Jimenez, J., Sobieszczyk, M. E., Hammer, S. M., Copenhaver, D. J., & Kulkarni, R. P. (2009). Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. *International Journal of Health Geographics, 8*(1), 49.

Chaplin, S. E. (2011). Indian cities, sanitation and the state: the politics of the failure to provide. *Environment and Urbanization, 23*(1), 57-70.

Chauhan, N. T., Prajapati, P., Trivedi, A. V., & Bhagyalaxmi, A. (2010). Epidemic investigation of the jaundice outbreak in Girdharnagar, Ahmedabad, Gujarat, India, 2008. *Indian Journal of Community Medicine, 35*(2), 294.

Chen, D., Cunningham, J., Moore, K., & Tian, J. (2011). Spatial and temporal aberration detection methods for disease outbreaks in syndromic surveillance systems. *Annals of GIS, 17*(4), 211-220.

Cinnamon, J., & Schuurman, N. (2010). Injury surveillance in low-resource settings using Geospatial and Social Web technologies. *International Journal of Health Geographics, 9*(1), 25.

Collier, P., & Dollar, D. (2001). Can the world cut poverty in half? How policy reform and effective aid can meet international development goals. *World Development, 29*(11), 1787-1802.

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *American Journal of Preventive Medicine, 32*(5), S112-S118.

Déglise, C., Suggs, L. S., & Odermatt, P. (2012). SMS for disease control in developing countries: a systematic review of mobile health applications. *Journal of Telemedicine and Telecare, 18*(5), 273-281.

DeSouza, S. I., Rashmi, M., Vasanthi, A. P., Joseph, S. M., & Rodrigues, R. (2014). Mobile Phones: The Next Step towards Healthcare Delivery in Rural India? *PloS One, 9*(8), e104895.

Doron, A., & Jeffrey, R. (2013). *The great Indian phone book how the cheap cell phone changes business, politics, and daily life*. Cambridge: Harvard University Press.

Duggal, R., & Gangolli, L. V. (2005). Introduction to review of healthcare in India. In L. V. D. Gangolli, R; Shukla, A. (Ed.), *Review of Healthcare in India*. Mumbai: Satam Udyog.

Elias, J., Harmsen, D., Claus, H., Hellenbrand, W., Frosch, M., & Vogel, U. (2006). Spatiotemporal analysis of invasive meningococcal disease, Germany. *Emerging Infectious Diseases, 12*(11), 1689-1695.

Elliot, P., Wakefield, J. C., Best, N. G., & Briggs, D. (2000). *Spatial epidemiology: methods and applications*. Oxford: Oxford University Press.

Elliott, P., & Wartenberg, D. (2004). Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspectives, 112*(9), 998-1006.

Ensor, T., & Cooper, S. (2004). Overcoming barriers to health service access: influencing the demand side. *Health Policy and Planning, 19*(2), 69-79.

Estrin, D., & Sim, I. (2010). Open mHealth architecture: an engine for health care innovation. *Science (Washington), 330*(6005), 759-760.

Fiordelli, M., Diviani, N., & Schulz, P. J. (2013). Mapping mHealth research: a decade of evolution. *Journal of Medical Internet Research, 15*(5).

Frenk, J. (2009). Reinventing primary health care: the need for systems integration. *The Lancet, 374*(9684), 170-173.

Gao, S., Mioc, D., Anton, F., Yi, X., & Coleman, D. J. (2008). Online GIS services for mapping and sharing disease information. *International Journal of Health Geographics, 7*(1), 8.

Garg, S., Singh, R., & Grover, M. (2012). India's health workforce: Current status and the way forward. *National Medical Journal of India, 25*(2), 111.

German, R. R., Lee, L., Horan, J., Milstein, R., Pertowski, C., & Waller, M. (2001). Updated guidelines for evaluating public health surveillance systems. *Morbidity and Mortality Weekly Report*, *50*(RR-13), 1-35.

Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis, 24*(3), 189-206.

Ghosh, A. (1996). Health Care and Globalisation: Case for a Selective Approach. *Economic and Political Weekly, 31*(8), 441-442.

Gupta, R., & Jain, K. (2012). Diffusion of mobile telephony in India: An empirical study. *Technological Forecasting and Social Change, 79*(4), 709-715.

Haberer, J. E., Kiwanuka, J., Nansera, D., Wilson, I. B., & Bangsberg, D. R. (2010). Challenges in using mobile phones for collection of antiretroviral therapy adherence data in a resource-limited setting. *AIDS and Behavior, 14*(6), 1294-1301.

Hashimoto, S., Murakami, Y., Taniguchi, K., & Nagai, M. (2000). Detection of epidemics in their early stage through infectious disease surveillance. *International Journal of Epidemiology, 29*(5), 905-910.

Heathcote, J., Elewaut, A., Fedail, S., Gangl, A., Hamid, S., Shah, M., . . . Krabshuis, J. (2004). Management of acute viral hepatitis. *World Gastroenterology News, 9*(1 suppl), 1-12.

Herma, D. C. (2013, February 6th, 2013). *Water Supply and Sanitation in Ahmedabad City.* Paper presented at the Preparing for the Urban Challenges of the 21st Century, New Delhi, India.

India, G. o. (1983). National health policy issued by the ministry of health and family welfare, government of India, 1982. *The Nursing Journal of India*, *74*(5), 121-128.

Istepanian, R. S., Jovanov, E., & Zhang, Y. (2004). Guest editorial introduction to the special section on m-health: Beyond seamless mobility and global wireless health-care connectivity. *Institute of Electrical and Electronics Engineers Transactions on Information Technology in Biomedicine, 8*(4), 405-414.

Istepanian, R. S., & Lacal, J. C. (2003). *Emerging mobile communication technologies for health: some imperative notes on m-health.* Paper presented at the Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2003., Cancun, Mexico.

Istepanian, R. S., Laxminarayan, S., & Pattichis, C. S. (2006). *M-health*. New York: Springer.

Jajosky, R., & Groseclose, S. (2004). Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. *BMC Public Health, 4*(1), 29.

Jamison, D. T., Breman, J. G., Measham, A. R., Alleyne, G., Claeson, M., Evans, D. B., . . . Musgrove, P. (2006). *Disease control priorities in developing countries*. New York: World Bank Publications.

Jamison, D. T., & Mosley, W. H. (1991). Disease control priorities in developing countries: health policy responses to epidemiological change. *American Journal of Public Health, 81*(1), 15-22.

Jones, S. G., & Kulldorff, M. (2012). Influence of spatial resolution on space-time disease cluster detection. *PLoS One, 7*(10), e48036.

Joumard, I., & Kumar, A. (2015). Improving Health Outcomes and Health Care in India. *OECD Economics Department Working Papers, 1184*.

Kamanga, A., Moono, P., Stresman, G., Mharakurwa, S., & Shiff, C. (2010). Rural health centres, communities and malaria case detection in Zambia using mobile telephones: a means to detect potential reservoirs of infection in unstable transmission conditions. *Malaria Journal, 9*(1), 96.

Kant, L., & Krishnan, S. K. (2010). Information and communication technology in disease surveillance, India: a case study. *BMC Public Health, 10*(Suppl 1), S11.

Kitron, U. (1998). Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. *Journal of Medical Entomology, 35*(4), 435-445.

Kraak, V. I., Harrigan, P. B., Lawrence, M., Harrison, P. J., Jackson, M. A., & Swinburn, B. (2012). Balancing the benefits and risks of public–private partnerships to address the global double burden of malnutrition. *Public Health Nutrition, 15*(03), 503-517.

Krause, G., Altmann, D., Faensen, D., Porten, K., Benzler, J., Pfoch, T., . . . Claus, H. (2007). SurvNet electronic surveillance system for infectious disease outbreaks, Germany. *Emerging Infectious Diseases, 13*(10), 1548.

Krishna, S., Boren, S. A., & Balas, E. A. (2009). Healthcare via cell phones: a systematic review. *Telemedicine and e-Health, 15*(3), 231-240.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics-Theory and methods, 26*(6), 1481-1496.

Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 164*(1), 61-72.

Kumar, R., Jaiswal, V., Tripathi, S., Kumar, A., & Idris, M. (2007). Inequity in health care delivery in India: the problem of rural medical practitioners. *Health Care Analysis, 15*(3), 223-233.

Kumar, S., Nilsen, W. J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., . . . Spruijt-Metz, D. (2013). Mobile health technology evaluation: the mHealth evidence workshop. *American Journal of Preventive Medicine, 45*(2), 228-236.

Labrique, A. B., Vasudevan, L., Kochi, E., Fabricant, R., & Mehl, G. (2013). mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. *Global Health: Science and Practice, 1*(2), 160-171.

Langmuir, A. D. (1963). The surveillance of communicable diseases of national importance. *The New England Journal of Medicine, 268*, 182.

Larson, J. S. (1999). The conceptualization of health. *Medical Care Research and Review, 56*(2), 123-136.

Li, J., Moore, N., Akter, S., Bleisten, S., & Ray, P. (2010). *mHealth for influenza pandemic surveillance in developing countries.* Paper presented at the 2010 43rd Hawaii International Conference on System Sciences, Honolulu, Hawaii.

Link workers' protest against Ahmedabad Municipal Corporation. (2014). *Times of India*. Retrieved from http://timesofindia.indiatimes.com/city/ahmedabad/Link-workers-protest-against-Ahmedabad-Municipal-Corporation/articleshow/30189576.cms

Manderson, L., Aagaard-Hansen, J., Allotey, P., Gyapong, M., & Sommerfeld, J. (2009). Social research on neglected diseases of poverty: continuing and emerging themes. *PLoS Neglected Tropical Diseases, 3*(2), e332.

Mandl, K. D., Overhage, J. M., Wagner, M. M., Lober, W. B., Sebastiani, P., Mostashari, F., . . . Koski, E. (2004). Implementing syndromic surveillance: a practical guide informed by the early experience. *Journal of the American Medical Informatics Association, 11*(2), 141-150.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research, 27*(2 Part 1), 209-220.

Martin, S. W., Michel, P., Middleton, D., Holt, J., & Wilson, J. (2004). Investigation of clusters of giardiasis using GIS and a spatial scan statistic. *International Journal of Health Geographics, 3*(1), 11.

McDonald, C. J., Schadow, G., Barnes, M., Dexter, P., Overhage, J. M., Mamlin, B., & McCoy, J. M. (2003). Open Source software in medical informatics—why, how and what. *International Journal of Medical Informatics, 69*(2), 175-184.

Mechael, P., Kwan, A., Hanzel, A., & Hedquist, C. (2014). Reshaping Health in a World of Mobile Phones. *Managing eHealth: From Vision to Reality*, 215.

Mechael, P. N. (2009). The case for mHealth in developing countries. *Innovations, 4*(1), 103-118.

Mehl, G., Vasudevan, L., Gonsalves, L., Berg, M., Seimon, T., Temmerman, M., & Labrique, A. (2014). Harnessing mHealth in Low-Resource Settings to Overcome Health System Constraints and Achieve Universal Access to Healthcare. In S. L. Lisa Marsch, and Jesse Dallery (Ed.), *Behavioral Health Care and Technology: Using Science-Based Innovations to Transform Practice* (pp. 239). Oxford, UK: Oxford University Press.

Morse, S. S. (2012). Public health surveillance and infectious disease detection. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science, 10*(1), 6-16.

Murray, C. J., & Lopez, A. D. (1997). Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *The Lancet, 349*(9063), 1436-1442.

Mykhalovskiy, E., & Weir, L. (2006). The Global Public Health Intelligence Network and early warning outbreak detection: a Canadian contribution to global public health. *Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique, 97*(1), 42-44.

National Informatics Centre. (2009). *Module A - Data Reporting and Analysis*. Retrieved from http://www.idsp.nic.in/idsp/usermanaul/ModuleA.pdf.

National Informatics Centre. (2010). *Manaual for Laboratory Diagnosis of Common Epidemic prone Diseases for District Public Health Laboratories*. Retrieved from http://idsp.nic.in/idsp/usermanaul/Laboratory_manual_IDSP.pdf.

Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide*: Cambridge University Press.

Nsubuga, P., White, M. E., Thacker, S. B., Anderson, M. A., Blount, S. B., Broome, C. V., . . . Sosnin, D. (2006). Public health surveillance: a tool for targeting and monitoring interventions. *Disease Control Priorities in Developing Countries, 2*, 997-1018.

Openshaw, S. (1984). The modifiable areal unit problem *Concepts and Techniques in Modern Geography No. 38*. Norwich, England: Geobooks.

Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis, 27*(4), 286-306.
Ostfeld, R. S., Glass, G. E., & Keesing, F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology and Evolution, 20*(6), 328-336.
Overhage, J. M., Grannis, S., & McDonald, C. J. (2008). A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. *American Journal of Public Health, 98*(2), 344.
Panackal, A. A., M'ikanatha, N. M., Tsui, F.-C., McMahon, J., Wagner, M. M., Dixon, B. W., . . . Morgan, J. (2002). Automatic electronic laboratory-based reporting of notifiable infectious diseases at a large health system. *Emerging Infectious Diseases, 8*(7), 685-691.
Patil, A. V., Somasundaram, K., & Goyal, R. (2002). Current health scenario in rural India. *Australian Journal of Rural Health, 10*(2), 129-135.
Patnaik, S., Brunskill, E., & Thies, W. (2009). *Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice.* Paper presented at the 3rd International Conference on Information and Communication Technologies and Development, Atlanta, Georgia.
Patnaik, U. (2007). Neoliberalism and rural poverty in India. *Economic and Political Weekly, 42*(30), 3132-3150.
Patrick, D. L., Bush, J., & Chen, M. M. (1973). Toward an operational definition of health. *Journal of Health and Social Behavior, 14*(1), 6-23.
Piarroux, R., Barrais, R., Faucher, B., Haus, R., Piarroux, M., Gaudart, J., . . . Raoult, D. (2011). Understanding the cholera epidemic, Haiti. *Emerging Infectious Diseases, 17*(7), 1161-1168.
Purcell, R. H., & Emerson, S. U. (2008). Hepatitis E: An emerging awareness of an old disease. *Journal of Hepatology, 48*(3), 494-503.
Rakodi, C. (2000). "Getting the pipe laid is one matter and getting the water flowing through the pipe is another": user views on public-sector urban water provision in Zimbabwe, Sri Lanka, Ghana and India. *International Planning Studies, 5*(3), 365-391.
Rao, M. (2005). The way the wind blows: Population policies in India. In L. V. D. Gangolli, R; Shukla, A. (Ed.), *Review of Healthcare in India*. Mumbai: Satam Udyog.
Reijn, E., Swaan, C. M., Kretzschmar, M. E., & van Steenbergen, J. E. (2011). Analysis of timeliness of infectious disease reporting in the Netherlands. *BMC Public Health, 11*(1), 409.
Robertson, C., Nelson, T. A., MacNab, Y. C., & Lawson, A. B. (2010). Review of methods for space–time disease surveillance. *Spatial and Spatio-temporal Epidemiology, 1*(2), 105-116.
Robertson, C., Sawford, K., Daniel, S. L., Nelson, T. A., & Stephen, C. (2010). Mobile phone–based infectious disease surveillance system, Sri Lanka. *Emerging Infectious Diseases, 16*(10), 1524.
Sachs, J. D., & McArthur, J. W. (2005). The millennium project: a plan for meeting the millennium development goals. *The Lancet, 365*(9456), 347-353.
Sahal, N., Reintjes, R., & Aro, A. R. (2009). Communicable diseases surveillance lessons learned from developed and developing countries: Literature review. *Scandinavian Journal of Public Health, 37*(2).
Saracci, R. (1997). The World Health Organisation needs to reconsider its definition of health. *The BMJ, 314*(7091), 1409.

Saravanan, V. S. (2013). Urbanizing diseases: contested institutional terrain of water-and vector-borne diseases in Ahmedabad, India. *Water International, 38*(7), 875-887.

Saravanan, V. S., Mavalankar, D., Kulkarni, S. P., Nussbaum, S., & Weigelt, M. (2015). Metabolized-Water Breeding Diseases in Urban India: Sociospatiality of Water Problems and Health Burden in Ahmedabad City. *Journal of Industrial Ecology, 19*(1), 93-103.

Sen, A. (1999). *Development as freedom*. New York: Oxford University Press.

Siegel, E. (2013). *Predictive analytics: The power to predict who will click, buy, lie, or die*. Hoboken: John Wiley & Sons.

Singh, A., Naik, V., Lal, S., Sengupta, R., Saxena, D., Singh, P., & Puri, A. (2011). *Improving the efficiency of healthcare delivery system in underdeveloped rural areas.* Paper presented at the 2011 Third International Conference on Communication Systems and Networks Bangalore, India.

Singh, S. (2012). Digital Divide in India: Measurement, Determinants and. *Digital Economy Innovations and Impacts on Society, 11*(20.1), 106.

Singh, S. K. (2008). The diffusion of mobile phones in India. *Telecommunications Policy, 32*(9), 642-651.

Slocum, T. A. (1999). *Thematic cartography and visualization*: Prentice hall Upper Saddle River, NJ.

Somani, A. (2012). *Slums and Incidence of Diseases in Ahmedabad*. International Development Research Centre-Industrial Technology Institute Working Paper.

Suresh, K. (2008). Integrated Diseases Surveillance Project (IDSP) through a consultant's lens. *Indian Journal of Public Health, 52*(3), 136-143.

Tandon, B., Gandhi, B., Joshi, Y., Irshad, M., & Gupta, H. (1985). Hepatitis virus non-A, non-B: the cause of a major public health problem in India. *Bulletin of the World Health Organization, 63*(5), 931.

Thacker, S. B., Choi, K., & Brachman, P. S. (1983). The surveillance of infectious diseases. *Jama, 249*(9), 1181-1185.

Tomlinson, M., Rotheram-Borus, M. J., Swartz, L., & Tsai, A. C. (2013). Scaling up mHealth: where is the evidence? *PLoS Medicine, 10*(2), e1001382.

Truelove, Y. (2011). (Re-) Conceptualizing water inequality in Delhi, India through a feminist political ecology framework. *Geoforum, 42*(2), 143-152.

UN-Habitat. (2004). The challenge of slums: global report on human settlements 2003. *Management of Environmental Quality: An International Journal, 15*(3), 337-338.

Unkel, S., Farrington, C., Garthwaite, P. H., Robertson, C., & Andrews, N. (2012). Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 175*(1), 49-82.

van Dijk, A., McGuinness, D., Rolland, E., & Moore, K. M. (2008). Can Telehealth Ontario respiratory call volume be used as a proxy for emergency department respiratory visit surveillance by public health? *Canadian Journal of Emergency Medicine, 10*(1), 18.

Vergu, E., Grais, R. F., Sarter, H., Fagot, J.-P., Lambert, B., Valleron, A.-J., & Flahault, A. (2006). Medication sales and syndromic surveillance, France. *Emerging Infectious Diseases, 12*(3), 416.

Verma, V., & Arankalle, V. A. (2009). Virological evaluation of domestic water purification devices commonly used in India emphasizes inadequate quality and need for virological standards. *Tropical Medicine and International Health, 14*(8), 885-891.

Vyas, S., Parikh, S., Kapoor, R., Patel, V., & Solanki, A. (2010). Investigation of an epidemic of hepatitis in Ahmedabad city. *National Journal of Community Medicine, 1*, 27-29.

Wagner, M. M., Tsui, F.-C., Espino, J. U., Dato, V. M., Sitting, D. F., Caruana, R. A., . . . Fridsma, D. B. (2001). The emerging science of very early detection of disease outbreaks. *Journal of Public Health Management and Practice, 7*(6), 51-59.

World Bank. (2012a). *India - Integrated Disease Surveillance Project (IDSP)*. Washington D.C.: World Bank Retrieved from http://documents.worldbank.org/curated/en/2012/11/17053254/india-integrated-disease-surveillance-project-idsp-india-integrated-disease-surveillance-project-idsp.

World Bank. (2012b). *Restructuring Paper On A Proposed Project Restructuring of Integrated Disease Surveillance Project*. Retrieved from Washington: http://documents.worldbank.org/curated/en/2012/03/16198953/india-integrated-disease-surveillance-project-restructuring-vol-1-2-main-report

World Health Organization. (1978). Alma Ata Declaration. *Geneva: World Health Organization*.

World Health Organization. (2006). *Communicable disease surveillance and response systems: A guide to planning*. Retrieved from Lyon: http://www.who.int/csr/resources/publications/surveillance/WHO_CDS_EPR_LYO_2006_2.pdf

World Health Organization. (2010). Public Health.

World Health Organization. (2014). International statistical classification of diseases and related health problems tenth revision (ICD-10). 2007. *World Health Organization: Geneva*.

Yach, D., Hawkes, C., Gould, C. L., & Hofman, K. J. (2004). The global burden of chronic diseases: overcoming impediments to prevention and control. *Jama, 291*(21), 2616-2622.

Yadav, K., Naik, V., Singh, A., Singh, P., Kumaraguru, P., & Chandra, U. (2010). *Challenges and novelties while using mobile phones as ICT devices for Indian masses: short paper.* Paper presented at the Proceedings of the 4th Association for Computing Machinery Workshop on Networked Systems for Developing Regions, San Francisco, California.

Yasnoff, W. A., O'Carroll, P. W., Koo, D., Linkins, R. W., & Kilbourne, E. M. (2000). Public health informatics: improving and transforming public health in the information age. *Journal of Public Health Management and Practice, 6*(6), 67-75.

Zhou, X., Bergquist, R., & Tanner, M. (2013). Elimination of tropical disease through surveillance and response. *Infectious Diseases of Poverty, 2*(1), 1.

Appendix A - Spatial Statistical Methods

A temporal/spatial-scan statistic is a well suited to waterborne disease surveillance (Martin et al., 2004). Using individual point data of reported cases, a specific cluster location and size can be identified with statistical confidence. Spatial-scan statistics individually evaluate every possible location with varying radii inside the surveillance extent to determine, for each location, if there is a higher than expected number of cases compared to the global distribution of cases (Kulldorff, 1997). In the absence of point data, the Getis-Ord GI* (GI*) method is a suitable alternative for hot spot analysis (Ord & Getis, 1995). This method aggregates cases by pre-defined polygons (e.g., ward or neighbourhood) and compares the local incidence rate of disease with the global rate to determine if each polygon has a comparatively high number of cases. The polygons used as the spatial unit for analysis should be as small as possible, or an appropriate size. Considering the risk factors in waterborne disease, an appropriate spatial unit would be a housing block. This is because spatial uncertainty in the location of risk factors can be minimized through narrowing down the possible “search area”. Case reports, therefore, should include precise location information of the patient to improve the spatial analysis in disease surveillance.

Appendix B – System Framework

Google App Engine instance of ODK Aggregate

First, the system administrator creates a Google account and a new project application from within Google App Engine assigning it a unique identifier and name. The ODK Aggregate installer is then run on a local computer with the user specifying the application identifier and Google account credentials. The installer then automatically builds and configures the code for ODK Aggregate to be run as the application on Google’s server and restricts access to authorized users for control within the application. Once the installer script is finished, the ODK Aggregate application can be accessed from the URL “https://application-id.appspot.com”.

Once data collection has begun, the data are streamed into a Google spreadsheet file attached to the main Google account. This is enabled by allowing the application from Google App Engine to have read/write access in Google Drive through the Drive API. The API is turned on from within the project’s configuration interface for the Google App Engine application. Once the API is turned on, the OAuth 2.0 protocol for authentication and authorization is used by ODK Aggregate and Google Drive to provide access to each other’s data. Specific credentials for API access, including a unique product key and application specific client id and email, are created. These are then uploaded into the site administration window of ODK Aggregate by the system administrator. Once the authorization process is complete, a new spreadsheet in the Drive folder of the associated Google account is created containing the results of the form. As long as publishing is enabled, and the OAuth 2.0 credentials remain valid, the spreadsheet in Google Drive is updated any time new data are collected.

Data form creation

The form was created in ODK Build, an online XML form-designer, which can directly push a completed form into ODK Aggregate. Build provides a GUI for creating new prompts and questions, allowing logical sequencing, specifying the data inputs and value range, and grouping questions into relevant themes. Once the form is completed, it is pushed directly into ODK Aggregate by specifying the URL, and then granting permission to ODK Build from the project’s Google account.

To prepare the form for data collection, the system administrator must make the form downloadable and be able to accept submissions from within the form management menu of ODK Aggregate. To prevent third parties from downloading the form or submitting data, anonymous data collection was disabled. Specific data collector accounts were created assigning a unique username and password for each individual Android device being used in the testing. Once connected, the data collector has access to the form to be completed, and is allowed to submit finalized forms into the ODK Aggregate database.

Form 1: Practice form used for field-testing with system

SURVEY TO BE ADMINISTERED TO INDIVIDUALS WITH CONFIRMED DIAGNOSIS OF HEPTATITIS A OR E ONLY.
આ સર્વે હીપેટાઇટિસ –A અને હીપેટાઇટિસ –B ના દર્દી માટે હાથ ધરવામાં આવશે.

1) Unique ID / યુનિક આઈ ડી:
2) Current Date / આજની તારીખ :
3) GPS Coordinates / GPS કોઓર્ડિનેટ્સ (રેખાંશ):
4) In the past month, have you experienced any of the following symptoms associated with Hepatitis A or E?
   પાછલા મહિનામાં તમે હીપેટાઇટિસ –A અને હીપેટાઇટિસ –B ના લક્ષણો (ચિહ્નો) તમે અનુભવ્યાં છે
   - Diarrhea / ઝાડા
   - Fever / તાવ
   - Vomiting / ઉલટી
   - Jaundice / કમળો
   - Abdominal Pain / પેટનો દુખાવો
   - Dark Urine / ઘાટા રંગનો પેશાબ
   - Loss of Appetite / ભૂખ મરી જવી (ભૂખ ના લાગવી)
   - Nausea / ઊબકા આવવા

   (IF YES, CONTINUE WITH SURVEY. IF NO, STOP SURVEY HERE)
   (જો જવાબ "હા" હોય તો આગળના પ્રશ્નો પૂછો. જો જવાબ "ના" હોય તો પ્રશ્નો પૂછવાનું બંધ કરો.)
5) Hepatitis A-E / હીપેટાઇટિસ A-E:
   a) Have you been diagnosed with Hepatitis A-E by a doctor?
      ANSWERS: No, Yes
      શું તમને હીપેટાઇટિસ –A-E ના રોગનું નિદાન થયેલ છે.
      જવાબ: ના, હા
      Government? / સરકારી?
      Private? / ખાનગી?
   b) When did you first come down with the associated symptoms (Fatigue, abdominal pain, dark urine or jaundice)?
      ક્યારે તમને પહેલી વાર આ રોગના લક્ષણો (ચિહ્નો) તમારી સામે આવ્યાં?
      (જેવાકે થાક લાગવો, પેટમાં દુખાવો, ઘાટા રંગનો પેશાબ, કમળો.)
   c) When were you diagnosed by a doctor?
      ક્યારે ડોક્ટર દ્વારા તમારા રોગનું નિદાન કરવામાં આવ્યું હતું?
   d) To your knowledge, has anyone in your household or village also been diagnosed or shown any symptoms?
      ANSWERS: No, Yes (symptoms), Yes (diagnosis)
      તમારી જાણ મુજબ તમારા ઘરના કોઈ પણ સભ્યો ને કોઈને પણ રોગનું નિદાન થયું છે? અથવા લક્ષણો (ચિહ્નો) દેખાયા છે?
      જવાબ: ના, હા (લક્ષણો માટે)
      જવાબ: ના, હા (નિદાન માટે)
      i) Are you able to identify the individual?
         શું તમે વ્યક્તિગત રીતે એને ઓળખી શકો છો?
6) Diagnosis Type? / નિદાનનો પ્રકાર?
7) Notes / નોધ

Creation of spatial files for KML

If the database is updated, a function is used to create a duplicate of the table, and then export the duplicate table as a CSV file to a local folder. Next, a python program was written to convert the database into a set of KML layers based on each case's disease status (e.g., confirmed hepatitis A, confirmed hepatitis E, confirmed both, and possible cases), as well as secondary indicators such as primary diet, and primary source of drinking water.

The python program uses the SimpleKML module to iteratively create the files. The home coordinates of the patient (latitude and longitude values obtained from the device's GPS) are used to represent the location of each point. A descriptive table for each patient is then created based on the other questions captured in the form including, for example, date of doctor diagnosis, type of symptoms, and type of doctor visited. This table of attributes describes their health profiles and will appear when each case is clicked from within the geovisualization interface. The date of diagnosis is converted into a timestamp element in the KML rendering, to allow for temporal filtering when opened.

Hot spot analysis

First, a grid-based shapefile was merged with population data provided from AMC (Ahmedabad Municipal Corporation, 2008). This shapefile was then imported into PostgreSQL with PostGIS. Next, a trigger function was written to create a duplicate table of the main case database as it gets updated. This table must be separate from the original data and table for export to CSV, as functions using PostGIS require spatial geometry data, which is missing in the original tables. This geometry column enables PostGIS to access the spatial information of each case directly from the table without having to geocode the data into a shapefile. Using the Well Known Text (WKT) spatial information standard, a point type geometry column is created by parsing the latitude and longitude information from the original form, using the WGS84 geographic coordinate system.

The next step in the analysis is to calculate the disease rates for each spatial unit, approximately one city block. A “JOIN” PostGIS function is written to sum the number of cases from the spatial table that fall within each kernel boundary. As per usual, the disease rate is calculated as the ratio of cases per 1,000 people. A new column is created in the shapefile from within PostgreSQL and the values are updated accordingly. A trigger function also listens for updates in the original database to re-calculate the sums and disease rates each time data are updated. Once the shapefile is updated from within PostgreSQL, its attributes are subsequently updated in the original shapefile.

Next, a python program using pysal, the python package of GeoDa, performs hot spot analysis on the shapefile. This script takes the case rate information and uses inverse distance weighting in the GI* equation (Equation 1) for analysis. It writes the results of the analysis to a

$$G_i^*(d) = \frac{\sum_j w_{i,j}(d) y_j - W_i^* \bar{y}}{s\{[(nS_{1i}^*) - (W_i^*)^2]/(n-1)\}(1/2)}, j = i$$

Equation 1. Getis-Ord GI* method for hot spot analysis

shapefile. If any kernels are found to have either High-High or High-Low statistically significant values, they are indicated within the attributes of the shapefile. A High-High value would indicate the unit is surrounded by areas with similarly high disease rates compared to the global rate of disease. A High-Low value would indicate the unit is surrounded by areas with lower disease rates. This resulting shapefile is automatically imported, and subsequently updated, into the PostgreSQL database. A trigger function then listens for updates in this shapefile and directly writes the updated file as a KML layer. The KML layer is colour coded to represent wards that are hot spots for the disease.

Geovisualization

The webpage has a division for the map interface, time-slider, layer selection menu, and attribute information windows. The map is enabled by calling the Google Maps V3 API. Several action buttons were created to allow the user to switch between satellite view, the default Google map style, and a customized map style. On loading, the map centres over Ahmedabad at a zoom level covering the extent of the city. The user can zoom in and out, but cannot pan into other geographic regions. It is hoped that this limiting prevents user-confusion.