Spring 6-3-2019

Closing the Loop: the Capacities and Constraints of Information and Communication Technologies for Development (ICT4D)

Phillip Nicholas Turman-Bryant
Portland State University

Let us know how access to this document benefits you.
Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds
Part of the Technology and Innovation Commons

Recommended Citation

10.15760/etd.6879

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
Closing the Loop: The Capacities and Constraints of Information and Communication Technologies for Development (ICT4D)

by

Phillip Nicholas Turman-Bryant

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Systems Science

Dissertation Committee:
Evan Thomas, Chair
Wayne Wakeland
Corey Nagel
Darrell Brown

Portland State University
2019
ABSTRACT

As a mechanism for collecting and sharing information, information and communications technologies (ICT) hold immense potential for individuals and institutions in low- and middle-income countries. Currently the distribution and adoption of ICTs—particularly mobile devices—has far outpaced the provision of other household services like clean water, sanitation, hygiene, or energy services. At the same time, the development and deployment of Internet of Things (IoT) devices including cellular- and satellite-connected sensors is facilitating more rapid feedback from remote regions where basic services are most limited. When used in conjunction with economic development or public health interventions, these devices and the feedback they provide can inform operation and maintenance activities for field staff and improve the monitoring and evaluation of outcomes for project stakeholders.

This dissertation includes three chapters written as journal articles. While each chapter is framed around the work and research efforts being undertaken by the Sustainable Water, Energy, and Environmental Technologies Lab (SweetLab) at Portland State University, the common thread that weaves all three investigations together is the theme of ICT-enabled programmatic feedback. The first chapter introduces the three theoretical lenses that inform these investigations and the ways that ICTs and the data they provide can (1) serve as more appropriate proxies for measuring access to services, (2) reduce information asymmetries between various stakeholders including communities, governments, implementers, and funders, and (3) enable more robust methodologies for measuring outcomes and impacts of interventions within complex adaptive systems. The second chapter presents a critical review of the methodologies and technologies being used to track progress on sanitation and hygiene development goals. Chapter three describes how simple sensors and weight measurements can be combined with complex machine learning algorithms to facilitate more reliable and
cost-effective latrine servicing in informal settlements. Chapter four presents the results from an investigation exploring how near-time feedback from sensors installed on motorized boreholes can improve water service delivery and drought resilience in arid regions of Northern Kenya. Finally, chapter five provides a summary of the three manuscripts and discusses the significance of this research for future investigations.
Dedicated to

Elizabeth, Oliver, and Abigail—my co-authors in life
ACKNOWLEDGEMENTS

As with any significant endeavor, this project would not have been possible without the support, mentoring, and encouragement of a whole constellation of individuals and mini-communities. I am first grateful to the ESUR IGERT program for making my time at Portland State possible. A huge thank you to Kim Heavener, Darrell Brown, Elise Granek, Heejun Chang, Veronica Dujon, David Ervin, Vivek Shandas, Alan Yeakley, Jennifer Allen, Randy Bluffstone, and Max Nielsen-Pincus. I am also grateful to my amazing IGERT colleagues, particularly my 2015 cohort. Thank you to Bruce Marron, Erin Looper, Erin Upton, Zuriel Rasmussen, and Amy Ehrhart as well as all the amazing IGERT fellows from previous cohorts that paved the way! A huge thanks as well to my friends and colleagues in the SweetLab and SweetSense, including Taylor Sharpe, Chantal Iribagiza, Emily Bedell, Lauren Stover, Christian Muragijimana, Danny Wilson, Jeremy Coyle, Skot Croshere, Katie Fankhauser, Zdenek Zumr, Zak White, Dexter Gauntlett, Kwasi Boateng, and Styvers Kathuni.

I am also grateful for the partners that made this research possible, including Sanergy in Nairobi, Kenya, the county and sub-county water officers in Turkana, individuals at Catholic Relief Services, the Turkana Water Project, and the Millennium Water Alliance (particular thanks to Alred Ireri for being an amazing host during my visit to Turkana). I am also grateful for the funding agencies that made this research possible, including the National Science Foundation, the Link Foundation, and the Sylff Association.

I am also indebted to my dissertation and comprehensive exam committees and mentors in this work: Arne Jacobson, Darrell Brown, Corey Nagel, and Wayne Wake-land. Particular thanks to Evan Thomas: for modelling unparalleled productivity, expansive generosity, and unwavering commitment to impact.

Finally, thanks to my different families. For the New Napkins and their support
through this academic odyssey. To my parents and brothers for their on-going love and support. To our extended family in Springwater for holding our family with tremendous care, candor, and consideration during our time in Portland. And to the love of my life and kiddos, to whom this project is dedicated. If it’s true that we are what we love, then I am grateful that I am becoming more like each of you everyday.
Table of Contents

Abstract i
Dedication iii
Acknowledgements iv
List of Tables ix
List of Figures x

Chapter 1 Introduction 1
1.1 Three Theoretical Lenses 2
1.1.1 Dynamic, In-Situ Monitoring as a Proxy for Use 5
1.1.2 Information Asymmetry as a Market Failure 12
1.1.3 Development as an Emergent Property of a Complex Adaptive Process 16
1.2 Summary of Chapters 23

References 25

Chapter 2 Measuring progress towards sanitation and hygiene targets: A critical review of monitoring methodologies and technologies 29
2.1 Introduction 31
2.2 Causal Modeling and Indicator Selection 32
2.3 Sanitation and Hygiene Service Ladders 35
2.4 Sanitation and Hygiene Indicators 38
2.4.1 Sanitation and Hygiene Beyond the Household 39
2.5 Monitoring Sanitation Outcomes 41
2.5.1 Accessible 41
2.5.2 Household Safety 42
2.5.3 Community safety 43
2.5.4 Use – Adherence 44
2.5.5 Equitable / Acceptable 46
2.6 Monitoring Hygiene Outcomes 47
List of Tables

3.1 Sample Characteristics ...................................................... 70
3.2 Performance metrics for the four prediction models, the actual implementation results, and a prediction model using low-use latrines. 78
4.1 Sample characteristics for all data and for the drought comparison sub-study. .......................................................... 102
4.2 Multilevel model results for all data. ..................................... 108
List of Figures

1.1 Above: Illustration of the open loop of development projects (used with permission). Below: Schematic of outputs, outcomes, and impacts in development projects. ................................................. 3
1.2 Closing the loop through sensors, novel financing mechanisms, and improved incentives (used with permission). ............................ 4
1.3 Example of outputs, outcomes, and impacts associated with different development technologies. .................................................. 7
1.4 Conceptual diagram of various monitoring methodologies and technologies as they relate to specific WASH interventions. Colors for each indicator are reflective of the indicator source (i.e., site, sensor, or person). ............................................................. 12
1.5 Diagram illustrating the lack of feedback between project implementers and project funders. ........................................................... 15
1.6 Diagram illustrating the lack of feedback between end-users and project implementers. ............................................................... 17
1.7 The coevolutionary process (Norgaard, 1994). ......................... 23

2.1 Diagram mapping the conditions and indicators for measuring impact in water, sanitation, and hygiene. ................................. 34
2.2 Mapping the categories of the sanitation and hygiene ladders to the desired outcomes. Line type designates the strength of the representation from each category to each outcome. .................. 36
2.3 Monitoring methodologies and technologies for sanitation outcomes. 43
2.4 Monitoring methodologies and technologies for hygiene outcomes. 48

3.1 Motion sensor installed in one of the latrines. ......................... 69
3.2 Chain of factors contributing to a latrine’s need to be serviced. .... 72
3.3 Relative importance of features used in the learner for predicting the probability of an overflow event for solid waste. The relative importance represented above is based on the mean decrease in Gini impurity from the randomForest learner. Gini impurity refers to the improvements in data classification that are contributed by each feature (Archer and Kimes, 2008). ................................. 75
3.4 Area under the receiver operating characteristic (AUROC) curve for solid (left) and liquid (right) waste overflow predictions. ........... 79
3.5 Sensitivity (Sens), specificity (Spec), negative predictive value (NPV), and positive predictive value (PPV) for solid waste overflow predictions over a range of probability thresholds. ........................................ 80

3.6 Average fill levels for the latrines with sensors (dashed line) and the latrines without sensors (solid line) for the baseline (pink) and intervention (blue) periods. .................................................... 81

4.1 Coupled human and natural systems framework to model how perceptions, policies, and behaviors mediate the dynamic interactions between human and natural systems. ........................................ 98

4.2 Average daily water system use for each county by year (line graph with left axis); annual precipitation (bar chart with right axis). .... 106

4.3 Left axis: average daily water system use for each county from 2017 through 2018. The red line indicates average daily use across all boreholes for each county, whereas the blue and pink dots show the individual observations for strategic and non-strategic boreholes, respectively. Right axis: the blue line describes the average daily precipitation across all pump sites for each county. The dark grey region demarcates the primary rainy season (March through May) and the light grey region demarcates the secondary rainy season (September through November). ...................................... 107

4.4 A: Average marginal effect of strategic borehole status on hours of utilization for each county. B: Average marginal effect of strategic borehole status on hours of utilization for the 2017 and 2018 drought seasons. ......................................................... 109

5.1 Conceptual diagram of the sources, measures, and outputs of monitoring and evaluation methodologies. Each research effort is framed in red. ................................................................. 124

5.2 Monitoring and evaluation schema applied to Sanergy research effort. Dashed lines represent feedback that is incorporated for adaptive project design and management. ............................... 127

5.3 Monitoring and evaluation schema applied to Kenya RAPID research effort. ........................................................................ 128

5.4 Phases of research in a translational science schema. T1TR and T2TR stand for Type 1 and Type 2 Translational Research (ICTR, 2019). 137

5.5 “Schematic illustration of the utilization of randomized controlled trial data and real world data through the lifecycle of a medical intervention” (Katkade et al., 2018). ................. 138

5.6 “General hierarchy of study types” (Katkade et al., 2018). ............... 140

5.7 “Hybrid effectiveness-implementation designs as part of the research continuum” (Cully et al., 2012). ................................. 141
Chapter 1

Introduction

The world’s most pressing problems do not fit discretely within one academic discipline, and international development is no exception. Low- and Middle-Income countries are currently undergoing a technological revolution, with the adoption of information and communication technologies (ICTs) like mobile phones that far outpace the provision of basic household services like clean water or electricity (UNCTAD, 2018). The following investigation explores the interface between ICTs and development objectives, what is often referred to ICT for Development (ICT4D). This research sits at the crossroads where Big Data and development collide, trying to translate the data deluge from ICTs into actionable knowledge that can improve people’s lives.

My work with the Sustainable Water, Energy, and Environmental Technologies Laboratory (SweetLab) at Portland State University has put me on the frontline of this research. With research collaborations that span the globe—including the London School of Tropical Medicine and Hygiene, UC Berkeley, Oregon Health and Sciences University, and Emory University—the SweetLab combines expertise in the areas of engineering, epidemiology, economics, public health, and public policy to explore the efficacy of development projects aimed at improving household health. In particular, the SweetLab manufactures cellular-enabled sensors that can be attached
to a wide variety of development devices, including cookstoves, water filters, water pumps, latrines, and hand-washing stations. These sensors provide a near-time data stream through cellular and satellite networks so that the efficacy of development interventions can be directly measured from the field. Over 1,500 of these sensors have been deployed in fifteen countries, and the SweetLab has been one of the technical leads in several randomized-controlled trials aimed at measuring the health impacts associated with cookstove, water filter, and latrine adoption for households in Rwanda, Kenya, India, and Bangladesh.

Through my work with the SweetLab, I explore how information and communication technologies like cellular- and satellite-enabled sensors can be integrated with development devices to improve development impact in three areas: the operation and maintenance of devices, the monitoring and evaluation of project outcomes, and project financing. A theme throughout this research is that data and devices do not operate in a vacuum but depend to a large degree on the complex social and institutional settings they are meant to support.

1.1 Three Theoretical Lenses

Anyone familiar with the work being done in the SweetLab has probably seen the image in Figure 1.1. It is the classic “open loop” of development that describes the way that information and resources flow from the West to the Rest with little feedback from the Rest. Funding from governments, NGOs, and individuals is used to promote various development devices like water filters, cookstoves, solar panels, improved latrines, or water pumps, but often very little is known about how end-users are impacted or the functionality of the devices. In an open loop scenario, projects are evaluated based on how effectively “outputs” are delivered with the available funding, with little consideration of the project “outcomes” and “impacts.”
The SweetLab theory of change is that this open loop can be closed by providing near-time feedback from sensors that are installed on devices to monitor use (see Figure 1.2). These sensors can provide important insights into end-user behavior and device functionality. This data can then be used to provide feedback to project managers to inform project design and implementation (monitoring of activities and outputs), to aid agencies to report progress on specific development goals (evaluation of outcomes), and to assess the overall impact of the project on the community.
of outcomes), and to investors wanting to support development projects through novel financing mechanisms like carbon or health credit markets (evaluation of impact). In this way, closing the loop can facilitate a virtuous cycle where the information being collected from the field becomes the primary enabler of project sustainability through improved maintenance, measurable performance, and impact-driven financing.

At least, that is how it is supposed to work.

The challenge lies in determining how much and how little can be learned from ICT data. A sensor installed on a pump, for example, doesn’t tell you who is using the pump. A solar home system installed in a home doesn’t tell you who is benefiting from the added light. Determining when and how sensor data can be used to improve
the outcomes and appropriately evaluate the impact of development projects is the primary focus of this investigation. In order to appreciate the capacities and constraints of ICT for development, it will be helpful to explore three theoretical lenses from sociology, economics, and complexity theory.

1.1.1 Dynamic, In-Situ Monitoring as a Proxy for Use

One of the central considerations in tracking progress on development projects and targets is the ability to define and measure access to household services. In their article, “A Theory of Access,” Ribot and Peluso redefine access as the “ability to derive benefit from things” (2003, pg. 153). Whereas traditional notions of access have relied on a conception of rights that are based on property relations, for Ribot and Peluso rights-based access mechanisms are but one strand of access that make up the general means, processes, and relations that determine who is able to gain, control, and maintain access to resources (2003, pg. 155). Other mechanisms of access include illegal access and structural and relational access mechanisms. Structural and relational access mechanisms can include technology, capital, markets, labor, knowledge, authority, social identity, and social relations. Taken together, these access mechanisms constitute the “material, cultural, and political-economic strands within the ‘bundles’ and ‘webs’ of powers that configure resource access” (2003, pg. 154).

By developing an analysis of access in which property is just one facet, Ribot and Peluso have presented a much richer and more complex conception of power that is characterized by a person or institution’s connectivity to the strands, bundles, and webs that configure resource access.

Access to electricity through the electrical grid is a good example of a benefit that could be analyzed from the perspective of rights versus ability. For example, in traditional rural electrification efforts, access is generally portrayed as a binary:
Does the household have a legal connection to the electrical grid or not? However, this reductionist view of access ignores illicit connections as well as the other facets of access that influence that household’s ability to derive benefits from the resource (e.g., reliability of the grid connection; costs associated with connection, consumption, operation, or maintenance; allocation of energy within the household; geographic prioritization of grid infrastructure, Lee et al., 2016). An analysis based on ability maintains the complexity of the access configuration. Actors who do not possess a resource can still benefit (i.e., still have access to some degree). Actors who do possess a resource may benefit to different degrees depending on the power relations underlying the mechanism of access or the allocation of the resource.

Another example can be found in the Millennium Development Goal to halve the number of people without access to clean water by 2015. Although the United Nations claimed that it reached its goal five years ahead of schedule, the metric of success reflects more of a rights-based theory of access. For example, when water pumps or improved water sources were installed, access was immediately counted without consideration of how that access was being maintained after installation. Evidence unfortunately indicates a large majority of pumps fail within a few years of being installed, and some are never repaired (Foster, 2013). In this regard, access can vary in time, it can be lost or controlled, and it may require additional mechanisms of access to be maintained.

**Outputs vs. Impacts**

With the renewed emphasis on access in the Sustainable Development Goals, it is important to be clear what is meant by access. Is access the right to own a latrine, or is it the ability to derive benefit from the latrine? Does access include use and an actual reduction in exposure to fecal contamination, or is the mere presence of a
For example, Figure 1.3 provides four examples of development devices that have been promoted in emerging economies as mechanisms for decreasing disease burden and alleviating poverty. Historically, project outputs have served as hopeful proxies for project impacts: improved cookstoves are assumed to decrease indoor air pollution and reduce the incidence of respiratory disease. However, these hopeful expectations are often based on faulty assumptions (e.g., exclusive use of the improved cookstove, reduced emissions based on laboratory testing, or the assumption that cooking would take place outdoors).

In reaction to the optimistic assumption that impact will naturally occur as a re-
sult of programmatic inputs, development practice has flipped almost to the opposite extreme in the past decade. Drawing on the extensive experience and rigorous methods used in medicine and epidemiology, a much greater emphasis has been placed on experimentally controlled trials as the primary means of evaluating the impact of development interventions (Whittle, 2016). This form of ex-post evaluation requires a comparison against a counterfactual—a subset of the experiment population that is not exposed to the intervention—in order to determine whether measurable improvements can be attributed to the intervention (Pritchett, 2012). More recently, some researchers have questioned whether controlled trials are an appropriate measure of impact when the system being analyzed is complex (Whittle, 2016; Woolcock, 2013). For example, Pritchett and Sandefur argue that there is little external validity to randomized controlled trial (RCT) results for development projects, particularly in interventions that depend on human behavior (Pritchett and Sandefur, 2013).

Mixed Methods in Measurement

There may be other reasons to question controlled trials as the primary method for measuring progress in development. As argued by Vianne Adams in Metrics: What Counts in Global Health, measurement has always played a central role in global development, but some forms of measurement have been given more weight than others (2016). Historically, measurement methodologies were a central feature of colonial management by their respective empires. Following World War II, the unit of measure was shifted to the nation-state, even as multilateral institutions were formed in response to global trends and development agendas. While the methods and measurements employed in development continue to evolve (e.g., GDP per capita, Human Development Index, Millennium Development Goals, Sustainable Development Goals), the role of measuring has only increased in its importance and its influence on the
conception, execution, monitoring, and evaluation of global development strategies.

Although some celebrate the emphasis on numbers as neutral, apolitical mediums for measuring progress, others question the ways in which global development and health metrics are constructed and the consequences of their application. On a basic level, there is a recognition that numbers often speak more loudly in public discourse and policy decisions than the localized, contextual, or ethnographic knowledge produced through qualitative research. For example, Chambers contrasts the relative popularity of Rapid Rural Appraisal (RRA), which uses more quantitative methods, with Participatory Rural Assessment (PRA), which focuses more on local engagement, participation, and empowerment (Chambers, 1997). In this regard, while both qualitative and quantitative methods can be extractive and highly biased by etic assumptions and methodologies, it is generally assumed that qualitative measures are more suited to preserving context, nuance, history, and complexity. Moreover, the time and costs associated with obtaining numbers (e.g., through population surveys or a randomized controlled trial for an experiment) can often exceed the time and costs associated with more qualitative methods (e.g., PRA, though some qualitative data requires a significant amount of time to appreciate the diversity and complexity of local cultures and contexts, e.g., Dubash, 2002; Wilson, 1993). Still, depending on the scale of the investigation, it is often not reasonable to explore questions or measure results exclusively through qualitative methods. More recently, mixed-methods investigations attempt to offset the weakness of both approaches by incorporating both quantitative and qualitative measures. For example, semi-structured interviews can be conducted with individuals, households, or focus groups to reveal dynamics or assumptions that would not have been captured with the exclusive use of surveys.

The strengths and weakness of both quantitative and qualitative approaches can be seen in two examples of research that question the simplifying assumptions made
while conducting household-level research. In her article, “Daughters, Decisions, and Domination: An Empirical and Conceptual Critique of Household Strategies,” Diane Wolf suggests that “the concept of household strategies misrepresents intra-household behaviour, obscures intra-household stratification by gender and generation, and stifles the voices of the unempowered—usually females and the young” (1990, pg. 44). In addition to her more general critique of the ways that households and individuals within households are merged and discussed interchangeably, she rejects the paternalistic assumption that “that respondents do not fully understand what they’re doing, or do not understand it correctly,” that they are unable to articulate it, or that they are subconsciously motivated (Wolf, 1990, pg. 66). Instead of constructing narratives to explain individual behavior, she argues that researchers need to elicit explanations from the individuals themselves.

In his study of energy allocation in households with solar electric systems, Arne Jacobson reaches a similar conclusion. Calling into question economic models that conceive of households as unitary, bargaining, or cooperative collectives that allocate resources with Pareto efficiency, Jacobson suggests that household energy allocation is negotiated on a complex terrain that includes gender, elder-junior dynamics, “family wealth, the geography of the home, and technical issues associated with solar system use” (2004, pg. 204). As a result, it cannot be assumed that solar electric systems automatically benefit the more vulnerable members of a household, and policies aspiring toward both environmental and social benefit need to consider these complex dynamics in order to have their intended effect.

As argued by Adams, the benefits of mixed-methods are significant. In describing the importance of ethnography she states, ”[t]he goal is often to capture this complexity without reducing the phenomena observed to simple forms that can be counted in ways that make one case just like another, one specific event comparable to the next”
In this way, ethnography provides insight into the particular and specific without making inferences to a broader scale or population. Ethnography keeps its knowledge local whereas controlled trials offer the potential to make inferences on broader scales of time and geography.

**The Metric Depends on the Target**

Figure 1.4 presents a conceptual diagram of the different methodologies and technologies being used to measure outcomes in water, sanitation, and hygiene (WASH) interventions. As a first attempt at capturing the strengths and weaknesses of different WASH indicators, this figure is also meant to show that the linkages between observed outcomes and measurable impacts are often tenuous.

Given the conflation of outputs and impacts and the doubt cast on controlled trials, what can be said about effective monitoring and evaluation of development projects? One of my central assumptions in approaching this research is that the choice of an appropriate measure depends on the desired target. Thus, in the international effort to promote access to clean water and improved sanitation, it is imperative that aid agencies clarify what is meant by access and use appropriate methodologies and technologies to measure progress. In this regard, electronic sensors may be a useful tool for dynamically monitoring use of development devices in the field. However, sensors on their own are insufficient as a means of gauging other aspects of access like equity within the household, quality of service, or adherence of use. Either way, in-situ sensors represent a significant improvement over the traditional metrics of coverage and self-reported use. Whether their data can serve as an appropriate proxy for impact, however, remains to be seen.
Figure 1.4: Conceptual diagram of various monitoring methodologies and technologies as they relate to specific WASH interventions. Colors for each indicator are reflective of the indicator source (i.e., site, sensor, or person).

1.1.2 Information Asymmetry as a Market Failure

Another important consideration in monitoring and evaluation is the extent to which information can flow easily between the interested parties. In their book on data-driven development interventions, Abhijit Banerjee and Esther Duflo find that the poor represented in their sixteen-country dataset “often lack critical pieces of informa-
tion and believe things that are not true” (2011, pg. 268). Looking at the examples of immunizing children, early-childhood education, fertilizer use, or HIV infection, they argue that the poor often face information constraints that impede their ability to make decisions that would contribute toward personal and societal welfare. At the same time, development interventions being implemented by western aid agencies lack the kind of end-user feedback that characterizes typical markets for consumer goods and services (Easterly, 2002). This lack of access to information can result in an information asymmetry between the West and the Rest that contributes to poor project outcomes or sustainability when there is little demand for the product or service being promoted. However, the lack of feedback between project implementers and project funders can also contribute to an information asymmetry that distorts incentives and reward outputs over impacts in foreign aid.

Information Asymmetry Defined

Information asymmetry was first explored by George Akerlof (1970) in his seminal paper, “The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism.” Using used car sales as a model for this theory, Akerlof demonstrates how the asymmetric distribution of information in the market—where buyers know more than sellers about the quality of the goods being sold—can encourage the sale of low-quality goods and engender consumer distrust. Therefore, in markets where it is difficult to distinguish between high and low quality goods, uncertainty about product quality artificially suppresses the buyer’s willingness to pay, which makes it more difficult for high quality products to compete in an information-constrained market.

Given that information asymmetry in a market results in the overall degradation of value of those goods, there are a variety of mechanisms that attempt to correct for this imbalance. For example, Michael Spence first examined how information
asymmetry in the job market is often balanced through an applicant’s use of signals: “observable characteristics attached to the individual that are subject to manipulation” by the applicant and that are typically costly to send (e.g., having a college education as proof of competency or diligence) (Spence, 1973). Similarly, Joseph Stiglitz (1975) first developed the theory of screening whereby information asymmetry is balanced when uninformed parties can induce the informed parties to share that information (e.g., insurance companies that offer lower premiums for different risk classes). More broadly, Stiglitz and Greenwald argued that “economies in which there are incomplete markets and imperfect information are not, in general, Pareto efficient” (Greenwald and Stiglitz, 1986, pg. 230). In other words, in markets where information is asymmetrically distributed, the invisible hand of the market does not provide the most efficient allocation of resources, and some form of intervention (e.g., government policies, taxes, or subsidies) can be used to more effectively distribute information to market stakeholders.

Information Asymmetry Within the West

As described by Easterly, aid agencies in the West often operate within a funding and accountability environment that rewards visible outputs and fidelity to proposals over actual impact (Easterly, 2002). As a result, aid work can suffer from the same information asymmetry described by Akerlof in the used car market: donors and investors (the buyers) have a hard time distinguishing between the higher and lower quality projects being proposed by aid agencies (the sellers), leading to an overall degradation of the quality of aid projects. The lack of clear signals of quality among aid projects can engender a general distrust of foreign aid, and “the industrial structure of foreign aid [can limit] competitive pressure on aid agencies” in three distinct ways (Easterly, 2002, pg. 5). First, the emphasis on coordination and collaboration
of aid agencies allows them to function as a “cartel of good intentions” that can exclude new competitors and limit pressures to reduce high costs. Second, the presentation of collaborating agencies as a “united front” makes it hard for donors and investors to differentiate the most effective channels for funneling resources. Finally, governments receiving aid are often forced to accept the projects being promoted by the aid cartel with little opportunity to choose between aid agencies or negotiate terms (2002, pg. 26). In this way, a focus on avoiding visible failures, an emphasis on short-lived outcomes, and a general pressure to maintain the status quo of aid disbursement results in a market failure that favors reports, frameworks, and procedures over sustained positive impact.

**Information Asymmetry Between the West and the Rest**

However, providing better feedback between aid agencies and donors would only solve half the problem of foreign aid dysfunction. According to Easterly,
“[t]he most obvious cause of bureaucratic dysfunction is that bureaucracies do not face the same degree of market feedback on their actions as private firms. This problem is particularly acute in foreign aid, where the intended beneficiaries have little ability to give feedback, and even when they do there is no mechanism by which the aid agencies suffer adverse consequences if the ‘customers’ are dissatisfied.” (2002, pg. 29)

In other words, development projects are prone to failure when there is little incentive to incorporate and respond to the feedback from end-users. Given that end-users are rarely in the position to pay for the products or services being promoted, their perspective is often peripheral to the outputs and indicators prescribed by the funding agencies. In this way, the emphasis on ex-ante evaluation and observable indicators as accountability mechanisms can marginalize the feedback from end-users that could facilitate more effective outcomes. As described by Easterly, “[t]he feedback from experimentation and learning by doing that is essential to success in any activity is...mostly absent. This robs the aid community of its historical memory,” and its beneficiaries of real progress (2002, pg. 33).

1.1.3 Development as an Emergent Property of a Complex Adaptive Process

Recently, there has been a push for broader recognition of development as an emergent property of complex adaptive economic and social system (Ramalingam, 2014; Easterly, 2015). Relating to Amartya Sen’s definition of development as the facilitation of choices, capabilities, and freedoms, Owen Barder argues that,

Development is not the sum of well-being of people in the economy, and we cannot bring it about simply by making enough people in the econ-
Development is instead a system-wide manifestation of the way that people, firms, technologies and institutions interact with each other within the economic, social and political system. Specifically, development is the capacity of those systems to provide self-organising complexity. Self-organising complexity in an adaptive system is never designed or deliberately built: it comes about from a process of adaptation and evolution. It follows that if we want to accelerate and shape development, we should focus especially on how the environment can be made most conducive for self-organising complexity to evolve. (2012)

In addition to the growing recognition that development is a complex adaptive process that cannot be reduced to the sum of specific interventions, there has been a greater emphasis on systems thinking as applied to development. For example, the Sustainable WASH Systems Learning Partnership was awarded a large USAID
contract specifically to explore how systems-based approaches can improve the sustainability of water, sanitation, and hygiene services. Similarly, the International Reference Centre for Community Water Supply and Sanitation (IRC WASH) recently hosted a symposium for practitioners and researchers focused on how to apply systems approaches to WASH service delivery.

**Complicated vs. Complex**

Although there is a growing consensus that development initiatives are complex rather than complicated, it is not always clear how complex and complicated processes are differentiated (Whittle, 2016). There seem to be few formal definitions of complex systems in the literature, with most descriptions relying on anecdotal analogies rather than formal criteria. For example, Amadei suggests that there are three types of problems: simple (where we “know the knowns”), complicated (where we “know the unknowns”), and complex (where we “don’t know the unknowns”; Amadei 2015, pg. x). Similarly, Norgaard notes that “[t]he logic is that development has gone awry because people have not been fully in control. Modernity is failing because, with control over nature only partial and with the social systems not fully rationalized, the project is like a poorly designed space mission, doomed to failure” (1994, pg. 12). According to Norgaard, sustainable development cannot be conceived as a space mission—a complicated problem wherein all the unknowns are known and the uncertainty can be quantified. Instead, he suggests that sustainable development is a complex problem with complex processes. As such, a different set of tools and assumptions are required.

One of the consequences of complexity is the inability to predict how changes or alterations in the system structure will change system behavior. Thus, attempts at modifying one aspect of the system often result in unintended consequences. As described by Norgaard, ”[e]nvironmental problems are problems of social organization.
This fact is frequently portrayed as a problem of our inability to devise social systems which can foresee and control the problems of new technologies” (1994, pg. 15).

One reason that prediction is almost impossible with complex systems is due to the interconnectedness of system components and the inability to predict how change will ripple through the chain of causality. Amadei captures this well when he characterizes well-being not as an isolated measure of health or income, but as a “property that emerges from the multiple functional or nonfunctional interactions in the community systems (social, environmental, infrastructure, economic and finance, political, and health at the individual, household, neighborhood, city, and regional scales)” (2015, pg. 5). As he describes it, there are two principle problems with reductionist and deterministic design and planning: first, that it neglects the dynamic feedback mechanisms between community issues, and second, that there could be common root causes to community issues that could be addressed in a more integrated and effective way (2015, pg. 6). Although this bears a sharp resemblance to the differential diagnosis advocated by Jeffrey Sachs in his book, The End of Poverty: Economic Possibilities for Our Time, Sachs has been criticized for his belief in rapid progress on poverty alleviation (i.e., ending extreme world poverty by 2025).

For some, the pace of change is vital for determining the consequences of change in complex systems. Even before a theory of complexity had been developed, Karl Polanyi recognized that the economic liberalization of the 19th Century had ramifications that continued to unfold well into the 20th Century, despite a strong shift toward protectionist policies and the repeal of the gold standard for currency regulation. As he describes it, the “secular tenets of social organization embracing the whole civilized world are not dislodged by the events of a decade” (2001, pg. 149). Norgaard comes to a similar conclusion when he states that “[u]nsustainable development, in this context, has resulted from technology outpacing changes in social organization.
Incentives and regulations must evolve with technologies” (1994, pg. 15).

The Philosophical Underpinnings of Progress

In his book, Development Betrayed: the end of progress and a coevolutionary revisioning of the future, Norgaard provides a novel explanation for the environmental degradation experienced in the past century. While a great many explanations have been put forth (e.g., natural scientists blaming expanding human population, economists blaming externalities, and others blaming the rapid introduction of new substances and pollutants), Norgaard suggests that the roots of environmental degradation are more philosophical: “the key premises of Western patterns of thinking help explain the cultural and biological destruction associated with modernism” (1994, pg. 61). In particular, he argues that the five philosophical underpinnings of modernity, “[a]tomism, mechanism, universalism, objectivism, and monism are not poor philosophical suppositions from which to reason... The problem is that these beliefs are embedded in our public discourse to the exclusion of other metaphysical and epistemological premises which are more appropriate for understanding the complexities of environmental systems and which are more supportive of cultural pluralism” (1994, pg. 63). Although these suppositions have been the source of astounding productivity over the past two centuries, he suggests that they are not suitable for thinking about complex systems, and in particular systems that include people.

Even though the five philosophical suppositions of Western rationality do not figure prominently as “the basis of thought and action for individuals, families, and small groups,” Norgaard suggests that they are the dominant assumptions in public discourse (1994, pg. 70). For example, problems that cannot be broken into their constituent parts (atomism) are often weeded out of public discourse, and objectivism is revealed in our implicit belief that we can isolate “the reality on which we derive our
knowledge from the reality on which we act” (1994, pg. 72). In contrast Norgaard 
notes that action in complex systems “changes the nature of parts and relations, 
typically in an irreversible manner” (1994, pg. 72). Furthermore, the introduction of 
new parts (e.g., agrichemicals, industrial wastes, etc.) creates brand new relations 
within the system. Similarly, our mechanistic assumptions lead to a belief that it 
is “possible to predict how systems will respond to different stimuli and hence to 
make them behave as we see fit” (1994, pg. 70). However, our interventions do 
not lead to new equilibria (a common assumption in neoclassical economics) but 
rather new problems with new relationships that continue to evolve. Finally, the 
universalist assumption can promote management from afar, centralization, and large-

scale factory-like operations, while our belief in monism promotes a cultural narrowing 
that rejects answers or indigenous knowledge that may be just as reliable (1994, 
pg. 72-73).

In contrast, Norgaard provides five alternate premises that are more appropriate 
for complex processes: holism instead of atomism, organic / evolutionary process 
instead of mechanism, contextualism instead of universalism, subjectivism instead 
of objectivism, and pluralism instead of monism (1994, pg. 62). It is important to 
note that Norgaard does not provide a formal distinction between complex versus 
complicated systems. In this regard, Norgaard is not trying to create a new dogma 
of philosophical suppositions that should be applied to all complex processes. To 
the contrary, Norgaard is generous in his praise of these premises and the productiv-
ity that they have wrought in the modern era. However, he cautions against their 
exclusive application in the public sphere where the positivist discourse has so thor-

oughly dominated in environmental and social policy-making. Rather than elevating 
the coevolutionary paradigm as the only means of interpreting and informing inter-
action with complex processes, he submits it as an equally valid (if not more valid)
A Coevolutionary Process

In response to the conception of development as a complicated process, Norgaard suggests that development should be conceived as a coevolutionary process. As he describes it, “[i]t is not to deny that people directly intervene in and change the characteristics of environments. The coevolutionary perspective merely stresses the next step, how different states of the environment alter the selective pressure and hence the relative dominance of species and relationships between species thereafter” (1994, pg. 36). In a coevolutionary process, we cannot predict how our attempts at improvement will affect the overall system: “There is no way that societies could keep track of all the flows that are quantifiable, no way that they could make sense out of them if they did, and no way to keep track of the unquantifiable flows at all. Sustainable development cannot be defined operationally” (Norgaard, 1994, pg. 20). As a result, Norgaard suggests that experimentation should be undertaken cautiously and on a small scale with as much monitoring of the evolutionary chain of events as possible. In the same way that Chambers suggests that development research should recognize the richness and validity of indigenous technical knowledge, a coevolutionary paradigm recognizes that communities began “developing” long before the industrial revolution or modernization. Whereas the development-as-progress paradigm suggests communities in the Global South just need better technology and more rational social organization, the coevolution-as-development paradigm recognizes that the community’s technology, knowledge, values, organization, and environment have been coevolving for hundreds, if not thousands of years.
1.2 Summary of Chapters

The three lenses of access, information asymmetry, and complex processes provide the theoretical foundation upon with this research is based. As described above, I am primarily focused on how in-situ, dynamic monitoring can facilitate better outcomes and impacts for development interventions. However, an exploration of sensor impact implicitly demands a theoretical framework that integrates social and technological systems. While much of the emphasis of this research is placed on the unique role played by sensors and the analysis of sensor data, it is important to acknowledge that technology does not operate in a vacuum. In addition to the technical analysis that this research has required (e.g., using machine learning to derive insights from simple data signals), I also analyze how these data insights are distilled, communicated, and operationalized.

The following represents a brief synopsis of the three research efforts that have been undertaken in fulfillment of this dissertation. Each research effort is framed by a particular project and publication related to work in the SweetLab. The second chapter provides an overview of the different methodologies and technologies that are
being used to track progress on sanitation and hygiene development goals. The third chapter investigates how data from motion sensors and weight measurements can be used to improve waste collection accountability and efficiency for a social enterprise in Nairobi, Kenya. The fourth chapter examines how near-time data from sensors on borehole pumps can be used to increase pump uptime and improve the management of water resources in five Northern Kenyan counties. Finally, the concluding chapter provides a summary of these three investigations and a discussion of implications for further research.
References


Chapter 2

Measuring progress towards sanitation and hygiene targets: A critical review of monitoring methodologies and technologies

This chapter was previously published as an article in the journal *Waterlines*.

Abstract

The Sustainable Development Goal (SDG) target for access to safe sanitation and hygiene represents a marked improvement over the target used during the Millennium Development Goal (MDG) period. The SDG target attempts to: explicitly address hygiene; eliminate inequalities within populations; evaluate sanitation services beyond the household; account for the accessibility, safety, acceptability, and affordability of service delivery; and improve the sustainability of services (WHO/UNICEF, 2015). However, the proposed indicators for monitoring progress in sanitation and hygiene still rely primarily on infrequent household surveys and census data. This paper provides a critical review of the sanitation and hygiene target and explores the potential gaps between the expanded understanding of access, the proposed monitoring strategies, and the desired impacts. A variety of innovative methodologies and technologies are reviewed, with specific attention given to their suitability for measuring and monitoring progress towards the sanitation and hygiene target.

Keywords: Sustainable Development Goals, sanitation, monitoring technologies, indicators, causal modeling
2.1 Introduction

Sustainable Development Goal (SDG) Target 6.2 is, “By 2030, achieve access to adequate and equitable sanitation and hygiene for all and end open defecation, paying special attention to the needs of women and girls and those in vulnerable situations” (WHO / UNICEF, 2015). While the corresponding MDG target emphasized a single outcome – access to improved sanitation facilities – the SDG sanitation target builds on this by incorporating adequacy and equity. By including “for all”, the target mandates that sanitation systems and services be available to all people at all times, regardless of age, gender, disability status or income level. Furthermore, the incorporation of child faeces disposal into the definition of open defaecation requires that all faeces be disposed of in a safe and hygienic manner, whether in an improved sanitation facility or treatment system. Lastly, the addition of special attention to women, girls and those in “vulnerable populations” requires that additional measures be met to provide for the special sanitation needs of women and girls, as well as to ensure that all people in “refugee camps, detention centers, mass gatherings, and pilgrimages” have adequate sanitation.

This expanded understanding of access recognizes that access should not be reduced to the binary measure of improved or unimproved facility type. Access can change over time, is influenced by structural and relational mechanisms, and operates on varying scales. All of these facets of access interact dynamically to influence the ability of individuals, households and communities to derive benefits from sanitation and hygiene services (Ribot and Peluso, 2003). Although this expanded understanding of access represents an improvement over the previous binary definition, a gap persists between the stated desire to improve service levels (e.g. to “promote progressive improvements in the quality of services based on the normative criteria of the
human right to water and sanitation i.e. accessibility, quantity, quality, acceptability, and affordability”) and the proposed monitoring strategies (WHO / UNICEF, 2015, pg. 7). For example, household surveys and spot checks are often the primary indicators for measuring progress, despite their being weak proxies for facility use and safety (Clasen, 2018). The combination of indicator uncertainty with an extended chain of inference (i.e. facility type and self-reported use as proxies for safety and actual use, and safety and actual use as proxies for health impact) makes monitoring progress on the SDG target difficult and contributes significantly to the overall uncertainty associated with evaluations of health impact.

After briefly illustrating how measurement uncertainty combines with inferential uncertainty in causal modeling, this paper then provides a critical review of the proposed sanitation and hygiene service ladders in relation to an expanded notion of access and service levels (i.e. accessibility, safety, use, equity and acceptability). Finally, a variety of innovative methodologies and technologies for monitoring are reviewed, with a discussion of their relative strengths and weakness.

2.2 Causal Modeling and Indicator Selection

In Evidence-Based Policy, Cartwright and Hardie describe how causal models are composed of a constellation of multi-faceted conditions that contribute to a desired effect (Cartwright and Hardie, 2012). While individual conditions are sufficient to cause an impact, they are considered unnecessary because of the variety of conditions that can produce the desired effect (e.g. improved sanitation, water or hygiene services). However, conditions are themselves composed of parts that are insufficient but non-redundant; that is, each part is unable to cause an impact on its own but it is a necessary component of the overall condition that contributes to the desired effect. Figure 2.1 provides a visual depiction of these insufficient but necessary compo-
nents of unnecessary but sufficient (INUS) conditions, where the sufficient conditions for health impact are represented as rings, and each ring segment represents a non-redundant component of that condition. This figure illustrates that all components of a condition must be monitored in order for the condition to be sufficient, and that two of the more common methods used for monitoring – spot checks and household surveys – serve as weak proxies for actual impact. For example, sanitation services will only have an impact on health if they are safe and if they are used. Similarly, sanitation services will only be used by all members of a household if they are acceptable, accessible, equitable and sustainable. As a result, a spot check can serve as a strong indicator of the accessibility of a facility, because it is visually verified, as opposed to household surveys that often rely on self-reporting to determine how often a facility is used, who uses it and whether it is acceptable. Similarly, facility type may be a weak proxy for the safety of the facility for the household, because there is no objective verification of how effectively the facility is separating fecal waste from human contact. Also, without any indicators for use or fecal sludge management, the assumption of health impact is dubious. However, uncertainty in inference can be reduced by using stronger monitoring indicators that are closer to the desired effect.

This point is echoed in a recent systematic review of indicator selection methods for water, sanitation and hygiene (WASH) monitoring, where Schwemlein et al. (2016) note that there is a general lack of consistency, specificity and relevancy in the indicators used by the projects and programmes included in their review. In particular, they suggest that better coordination of WASH indicators could help “identify weaknesses in data collection,” “inform decisions in WASH policy and practice,” and “facilitate comparison of projects, programs, and interventions” (2016, pg. 2). However, Schwemlein et al. argue that a more formal process for selecting indicators and organizing data collection is needed to improve transparency and improve coordina-
Figure 2.1: Diagram mapping the conditions and indicators for measuring impact in water, sanitation, and hygiene.
tion in WASH interventions. Notably, they recommend that the indicator-selection process should be explicitly tied to the outcomes of interest, based on the purpose and scope of the intervention. Finally, they suggest that proposed indicators should be evaluated using objective selection criteria, including whether the proposed indicator is measureable, reliable and sensitive to changes in the outcome of interest. They also argue that candidate indicators must be valid, that is, “[t]here must be an accurate correlation between an indicator and the issue for which it is supposed to proxy,” based on existing data (Schwemlein et al., 2016, pg. 11).

2.3 Sanitation and Hygiene Service Ladders

Recognizing that sanitation services can include a variety of levels, the Joint Monitoring Program (JMP) has updated its service ladder to define five thresholds. Like the service ladder used for the MDG sanitation target, three of the categories designate the type of sanitation facility: unimproved, limited or basic, where “limited” refers to latrines that are shared by two or more households. “Open defecation” describes the deposition of human feces directly in the environment, and “safely managed” designates a basic sanitation facility that is not shared and where excreta are safely disposed of in situ or treated off site. Similarly, the hygiene service ladder is primarily concerned with the presence of handwashing facilities on the premises, where basic facilities have soap and water and limited facilities have no soap or water. Figure 2.2 maps these categories from the proposed sanitation and hygiene service ladders to the outcomes identified in the target descriptions (WHO / UNICEF, 2015). Although a variety of outcomes are recognized in the target descriptions (i.e. accessibility, safety, equity and acceptability) and use is implied, many of these components are not explicitly represented in the proposed service ladders.

ACCESSIBLE: Accessibility is defined as “facilities that are close to home that
Figure 2.2: Mapping the categories of the sanitation and hygiene ladders to the desired outcomes. Line type designates the strength of the representation from each category to each outcome.
can be easily reached and used when needed” (WHO / UNICEF, 2015, pg. 11). Accessibility is well represented in both service ladders, with access being inferred directly from the observation of sanitation and hygiene facilities on the premises.

**SAFE:** The safety of the sanitation facility for the household – how well it separates excreta from human contact within the household – is not represented in the ladder. Household safety is indirectly inferred from the sanitation facility type, where basic sanitation facilities are assumed to adequately separate excreta from household contact, and unimproved sanitation facilities are not. The safety of the sanitation facility for the community – how well it separates excreta from human contact beyond the household – is better represented in the ladder. Like household safety, community safety is inferred indirectly from the sanitation facility type, but community safety is addressed directly by the “safely managed” category.

**USED:** Although sanitation and hygiene facility use is key for realizing health benefits, use is not addressed explicitly in any of the service categories. With respect to hygiene, it is also helpful to distinguish between adherence and technique, where technique designates efficacy in removing contamination, and adherence designates the consistency of use. Although none of the service categories address handwashing technique or adherence, regular and effective handwashing that coincides with sanitation behaviours is also important for realizing health benefits.

**EQUITABLE / ACCEPTABLE:** Equity is defined in the target as the “progressive reduction and elimination of inequalities between sub-groups” (WHO / UNICEF, 2015, pg. 11), but neither equity nor acceptability are represented explicitly in the sanitation ladder. While population-level inequalities could be inferred based on adoption rates or the elimination of inequalities across population sub-groups, more direct measures of equity and acceptability may be needed to capture intra-household use and the acceptability of specific sanitation and hygiene
interventions. For example, the needs of women and girls require special considera-
tion, given females’ higher risk of experiencing harassment and violence when safe and 
private sanitation facilities are not available (Bangdiwala et al., 2004; SHARE, 2014).
Beyond considerations of equity, acceptability could depend on the desirability of the 
technology and whether it is aspirational. In either case, if reliable measures of use 
are available, it is possible that equity and acceptability could be inferred indirectly 
from sanitation and hygiene facility use.

2.4 Sanitation and Hygiene Indicators

Although the inclusion of the “safely managed” category represents substantial progress 
in the evolution of the service ladder, there is still a great deal of ambiguity surround-
ing the indicators that will be used to monitor outcomes. Currently, the JMP plans 
on using household surveys and regulatory data as the main data sources for observ-
ing household sanitation facility types. Sanitation facility type and attributes will 
then be used to infer other outcomes like safety and use. For example, lack of use 
could be inferred directly from a non-functioning toilet, and unimproved sanitation 
facilities would be assumed to provide unsafe management both within and beyond 
the household (WHO / UNICEF, 2015).

Given that hygiene was not addressed in the MDG targets, its inclusion in SDG 
6.2 highlights the growing consensus that water, sanitation and hygiene are indelibly 
linked and cannot be treated in isolation. Hygiene is defined as “the conditions 
and practices that help maintain health and prevent disease including handwashing, 
menstrual hygiene management and food hygiene” (WHO / UNICEF, 2015, pg. 11). 
While food hygiene was identified as one of the top priorities for health and nutrition, 
it was ultimately determined to be outside the scope of WASH monitoring. Similarly, 
menstrual hygiene management is mentioned but is not addressed specifically in the
indicators. Thus, hygiene as described by the service ladder is primarily concerned with handwashing. Like the sanitation ladder, the hygiene service ladder uses the presence of a handwashing facility as a proxy for quality and use.

### 2.4.1 Sanitation and Hygiene Beyond the Household

One clear advance of the proposed indicators is the focus on fecal sludge management. The indicator defines “safely managed sanitation” as systems in which fecal waste is transported through a sewer to a designated location, is collected from systems by a process that limits human contact, and is transported to a designated location or undergoes, as a minimum, secondary treatment or “primary treatment with long ocean outfall for sewerage” or treated at a “managed disposal site” or wastewater treatment plant or “stored on site until...safe to handle and re-use” (WHO / UNICEF, 2015, pg. 28). This indicator is designed to encompass essential services and operational requirements for public health benefits (Feachem et al., 1983; Shuval, 2003; Escamilla et al., 2013). At the same time, the indicator does not evaluate the integrity of the system or services; neither is there consideration of sustainability, where sustainability is defined as the ability to sustain “services to ensure lasting benefits” by “safely managing human waste” (WHO / UNICEF, 2015, pg. 8). Also, given the higher infectivity of child faeces, special consideration should be given to their disposal and management (Rand et al., 2015; UNICEF, 2015).

The JMP indicators also acknowledge that monitoring the safe management of excreta requires a full fecal waste flow framework that spans the service chain from containment to reuse or disposal. While information about containment can be collected from household surveys, the JMP proposes to monitor the emptying, transport and treatment of fecal waste using a combination of utility, population and household data to estimate safe management through the service chain. As a result, “reuse and
disposal would not be monitored initially at a global level” (WHO / UNICEF, 2015). On-site treatment and disposal would be inferred based on a variety of factors, including the sanitation facility type, construction quality, frequency of use, population density, geographic conditions, and urban versus rural location (WHO / UNICEF, 2015). Off-site treatment will initially be estimated from utility records, based on the number of sewer connections and installed treatment facilities. Off-site treatment for excreta that are collected and transported from septic tanks and pit latrines could then be estimated using records from trucks disposing waste at wastewater treatment plants (WHO / UNICEF, 2015).

Monitoring sanitation and hygiene beyond the household also entails the evaluation of institutions such as schools and health clinics, where facility use and the risk of exposure to fecal pathogens are high. The negative impacts of incomplete sanitation coverage at the community level have been documented in field studies and systematic reviews (Barreto et al., 2007; Moraes et al., 2004; Geruso and Spears, 2018). A study of city-wide sanitation improvements in Salvador, Brazil, saw overall reductions in the prevalence of diarrhoea by 21%, and in high-risk areas with high baseline prevalence the reduction was 43% (Barreto et al., 2007). The new guidelines define three simplified service ladders (basic, unimproved and no service) for sanitation, handwashing and menstrual hygiene facilities. Institution-based indicators will rely on a combination of sector-based management information systems (MIS), site surveys of facilities, and household surveys that incorporate questions about extra-household facility access and use (WHO / UNICEF, 2015). In addition, the JMP acknowledges the need to monitor WASH services for vulnerable populations (e.g. refugee camps or detention centres), and data will primarily be sourced from surveys conducted by the United Nations High Commissioner for Refugees (UNCHR) and the International Committee for the Red Cross (ICRC) for global reporting (WHO /
2.5 Monitoring Sanitation Outcomes

The following section provides an overview of relevant practices and technologies for monitoring sanitation outcomes. As no one practice or technology is adequate for monitoring progress in sanitation, it is important to note that some practices and technologies are better suited for monitoring specific sanitation outcomes. Figure 2.3 provides a visual mapping of each methodology to each sanitation outcome.

2.5.1 Accessible

Household surveys and national censuses are the most common methodologies used for assessing a household’s access to sanitation facilities (Clasen et al., 2012). There are a variety of advantages to using surveys for evaluating access. First, as one of the most common tools for gathering household information, there is a growing knowledge base that facilitates comparison across time and geography. Second, appropriate survey design can result in higher validity and reliability of survey responses. Third, administering surveys in households allows for interaction with household members and direct observation of sanitation facilities. Thus, while survey questions can differentiate which members of the household are able to access a sanitation facility, direct observation allows an observer to verify the sanitation facility type, its functionality, whether it is private or shared, and its proximity to the household. Still, unless there are repeated visits to the household, census surveys and spot checks only provide a static measurement of the sanitation facility’s accessibility and functionality (Thomas and Mattson, 2013).
2.5.2 Household Safety

The type of sanitation facility and whether it is private are the two main proxies used to determine whether excreta are safely separated from human contact within the household. However, there are very few methods for directly measuring the quality of specific sanitation facilities. One exception is the Sanipath Rapid Assessment Tool created by researchers at the Center for Global Safe Water at Emory University in the USA. The Sanipath tool provides an assessment of exposure to fecal contamination by measuring the level of fecal contamination associated with different transmission pathways (e.g. drinking water, latrines, produce, open drains, etc.). These microbial loads are combined with surveys that characterize household behaviours to generate risk assessments for each exposure pathway. For example, a household may use a private pit latrine with a slab, but the Sanipath tool could be used to estimate the actual risk of exposure to fecal contaminates based on the level of contamination in the latrine and the behaviours of the users. While the Sanipath tool is primarily designed to evaluate the level of exposure to fecal contamination for an entire community, the methodology could be adapted to the household scale. The ability to combine microbial testing with survey responses is also a strength, as the surveys facilitate a more nuanced characterization of individual sanitation and hygiene practices. However, the tool depends on the ability of local laboratories to conduct testing in a sterile environment with sufficient equipment. Also, unless the Sanipath assessment is performed regularly, the measurement represents a snapshot in time that is not able to monitor changes in behaviour, fecal contamination in the environment, or the functionality of sanitation facilities (Sanipath, 2014).
2.5.3 Community safety

Similar to the situation for monitoring household safety, there are very few methodologies that have been developed to directly verify the safe management of excreta beyond the household. While safe management is often assumed for sanitation facilities that are connected to a sewer system, septic tank or pit latrine, the actual verification of waste removal, transport and treatment represents a significant challenge for monitoring community safety. Data from utilities could be used to estimate the safe treatment of excreta based on the number of household connections and the conveyance to installed treatment facilities. Similarly, records from disposal trucks could be used to estimate the number of households where waste is safely collected and removed (WHO / UNICEF, 2015). However, unless records from the point of collection to the point of treatment can be corroborated, utility and waste removal
estimates may underestimate leakage or the deposition of waste directly into the environment.

Although they may not provide an accurate measure of the level of exposure to fecal contamination in the environment, records from utilities and waste collectors can be used to verify that excreta are being collected and conveyed to treatment facilities. For example, Sanergy Inc. in Kenya has partnered with SweetSense Inc. to use motion sensors to optimize sanitation waste collection operations. The sensors are also able to send alerts from the latrine operator or the waste collector through Radio Frequency Identification (RFID) tags that are directly integrated into Salesforce, a logistics and customer management platform. Similarly, x-runner in Peru and the Water and Sanitation Program in India are able to track the installation and management of improved toilets through Near Field Communication (NFC) tags and Quick Response (QR) codes that are scanned and tracked through Salesforce and Open Data Kit, a mobile survey application (Robiarto et al., 2014; Nique and Smertnik, 2015).

2.5.4 Use – Adherence

Although not explicitly represented in the sanitation service ladder, sanitation facility use is a key indicator for measuring sanitation facility efficacy (Clasen et al., 2014). However, the verification of sanitation facility use and a household’s adherence to use is incredibly challenging.

BRAC in Bangladesh has developed a Qualitative Information System (QIS) that incorporates a combination of spot-check indicators with survey questions to assess latrine use. In a study comparing three latrine utilization methodologies, including surveys, observations, and motion detector sensors, there was a strong correlation between latrine spot-check indicators and BRAC’s QIS indicators. There was also a positive correlation between self-reported latrine use and sensor-recorded latrine use,
although self-reported use was significantly greater than sensor-recorded use. While households reported an average of 32.8 latrine uses over four days, sensors recorded an average of 21.7 uses, perhaps indicating recall or courtesy bias in self-reporting (Delea et al., 2017). Given the different scales used, no comparison was drawn between sensor-recorded use or self-reported use and the spot-check indicators.

Clasen et al. suggest that spot-check indicators and sanitary surveys “are subjective and may lack necessary sensitivity and specificity to quantify patterns of use” (Clasen et al., 2012). In an experiment comparing motion-detector-sensor-equipped latrine use against structured observations, they found that sensor-recorded use and observed use agreed within two latrine use events 93.9% of the time over 228 observation periods. They also found strong evidence of reactivity to structured observation, as the sensors recorded significantly more latrine events during observation periods compared with non-observation periods (Clasen et al., 2012). O’Reilly et al. also recorded a high level of agreement between sensor-reported events and structured observations (O’Reilly et al., 2015).

In a similar study, Sinha et al. found that mean reported “usual” daily use was almost twice the average daily sensor-recorded use (7.09 versus 3.62 events). While there was better agreement between reported use and sensor-recorded use from the previous 48 hours (4.61 versus 3.59 events), the predicted number of latrine events using the 48-hour recall measure was still 60% greater than the average number of events recorded by the sensors (Sinha et al., 2016).

In this regard, sensors like the passive latrine use monitor (PLUM), provide perhaps the best estimate of actual latrine utilization. They are subject to less reactivity compared with structured observations, can provide higher resolution data over longer observation periods in near time, and can be incorporated unobtrusively in a variety of sanitation facility settings. However, unlike structured observations and
self-reporting, PLUMs are not able to differentiate which individuals are using the sanitation facility in the household. Also, while PLUMs have a high degree of accuracy for household latrines, use of public latrines is more difficult to characterize due to shorter inter-arrival times. As a result, sensors like PLUMs should ideally be combined with surveys or observations to better characterize sanitation facility use and adherence.

2.5.5 Equitable / Acceptable

Although elimination of inequalities and the special needs of women and girls are addressed in the sanitation target, the JMP’s current proposal for measuring inequalities involves a comparison of population sub-groups that are disaggregated by “income, gender, age, race, ethnicity, migratory status, disability, geographic location, and other characteristics relevant in national contexts” (WHO / UNICEF, 2015). However, evidence shows that inequalities in sanitation facility use can occur at an intra-household scale as well as a societal scale (Jenkins and Curtis, 2005; Coffey et al., 2014). In addition, more nuanced methodologies may be needed to incorporate the specific needs of women and girls to ensure the acceptability, security and privacy of sanitation facilities. Given the sensitivity of sanitation subjects and the influences of cultural and religious norms, qualitative methodologies like ethnography and semi-structured interviews may be needed to accurately gauge acceptability and characterize intra-household sanitation behaviours. For example, O’Reilly found that ethnographic and motion detector data were highly complementary and useful for comparing sanitation practices between groups that differed in geography and religious affinity (O’Reilly et al., 2015).
2.6 Monitoring Hygiene Outcomes

In contrast to the challenges associated with monitoring progress on sanitation outcomes, the proposed methodology for monitoring progress on hygiene is relatively direct. Since 2009, the JMP has used the “observation of the place where household members wash their hands and the presence of water and soap” as the primary indicator of handwashing behaviour (WHO / UNICEF, 2015, pg. 21). As a result, the JMP is able to measure the hygiene service ladder directly through household surveys and extrapolate those estimates to the broader population base.

While the monitoring of hygiene facilities is relatively straightforward, the following section provides a summary of different practices and technologies that have been used to monitor specific hygiene outcomes. Actual handwashing behaviour is still challenging to monitor, but it is possible that the type of hygiene facility could serve as an adequate proxy for access and use for mixed-purpose large-population surveys (Ram, 2013). Figure 2.4 provides a visual mapping of each methodology to each hygiene outcome.

2.6.1 Accessible

As when evaluating sanitation, household surveys and censuses are the easiest indicators for evaluating access to hygiene facilities. Easily combined with spot-check indicators that facilitate direct observation of handwashing facilities and materials, rapid observations are used almost exclusively in large-population surveys where hygiene is one among many behaviours of interest. As a direct measure, rapid observations are cost-effective, efficient and more reliable than survey responses (Cairncross et al., 2005; Ram, 2013). However, verification of the handwashing facility does not provide information about individual hygiene practices within the household, whether hand-
Figure 2.4: Monitoring methodologies and technologies for hygiene outcomes.

washing is performed at critical times (e.g. after defaecation or before meals), or the efficacy of handwashing and its consistency across time.

2.6.2 Use – Technique

Measurement of microbial hand contamination through laboratory measurements or visual inspection are two methodologies that are used to verify handwashing efficacy. While research has shown a positive correlation between hand contamination and health outcomes (Luby et al., 2009; Pickering et al., 2010), measurement of hand contamination is relatively expensive, time-consuming, and may require access to a microbial laboratory facility (Ram, 2013). Observation of handwashing practice can be a useful method for verifying the use of soap, the duration of handwashing, and the method for drying, but respondent behaviour may be influenced by the presence of an observer (Sagerman et al., 2010). Visual inspections of hand contamination can be performed efficiently and are positively associated with microbial contamination and
observed handwashing (Pickering et al., 2010). However, a high inter-rater reliability is important for avoiding subjectivity bias between multiple enumerators (Ram, 2013). Also, as a static indicator, the measurement of hand contamination is not able to capture how quickly recontamination occurs after washing. For example, Ram et al. found a high level of recontamination within two hours of a thorough handwashing with soap (Ram et al., 2011).

2.6.3 Use – Adherence

Although handwashing with soap at critical times (e.g. after defaecation and before meals) has been identified as one of the most cost-effective behaviours for preventing infection, verification of handwashing adherence remains a challenge. Indicators like the presence of soap and water and handwashing efficacy are positively correlated, but it is still unclear how well these indicators predict handwashing behaviour (Ram, 2013). For example, Biran et al. (2008) found that only 2 out of 26 handwashing indicators used to classify households as “handwashing” – the presence of soap beside the latrine and soap in the yard – were significantly correlated with classifications of households based on structured observations.

Self-reported behaviour is one of the most common indicators used to assess hygienic practice. However, self-reported handwashing usually overestimates actual handwashing due to the social desirability associated with handwashing. For example, while 77% of respondents in a Bangladesh study reported handwashing with soap after defaecation, only 32% were observed to do so (icddr, 2008). When accounting for actual soap use, the discrepancy between reported and observed handwashing persists but decreases slightly (Ram, 2013).

Structured observation has typically been used as the gold standard for comparing different handwashing measures. Structured observations can record more detailed
information about how hands are washed, when hands are washed, and who washes their hands, but it is important that the timing and location of observations include as many members of the household as possible and critical events like meal preparation and consumption (Biran et al., 2008). Given that an observer’s presence has been shown to increase the number of handwashing events by as much as 35%, Ram et al. (2010) question whether structured observations should be the standard for comparison, due to high reactivity. Unlike self-reports, however, structured observation provides a more dynamic measure of handwashing behaviour over time. Repeated spot-checks also provide a more dynamic measure of handwashing behaviour. Webb et al. (2006) determined that six separate spot-checks are needed to reliably estimate a household’s hygiene practices, although repeated visits may also increase reactivity (Arnold et al., 2015).

While studies that monitor the consumption of cleansing products have been conducted in high-income countries, there are only a few examples of studies that have tracked soap purchases or soap weight differences as a proxy for handwashing behaviour (Ram, 2013). For example, Gadgil et al. (2010) found a positive correlation between consumption of bar soap and observed handwashing events. However, Luby et al. observed no differences in soap purchases between the treatment and control groups in a handwashing intervention, despite differences between the two groups in the presence of soap and water and handwashing techniques (Luby et al., 2009).

Sensors can also provide an objective and non-obtrusive characterization of handwashing behavior. For example, SmartSoap is an ordinary looking bar of soap with an embedded accelerometer that measures motion on three axes. On its own, SmartSoap can provide an accurate count of the number of times the soap bar is used each day, although it cannot differentiate users or determine whether handwashing coincides with critical events like defecation or meals. However, by combining SmartSoap data
with data from a motion sensor placed on the vessel holding water for anal cleansing, Biran et al. were able to detect handwashing events after defecation. Although overall soap-use increased, they found that there was no increase in the number of soap uses following defecation (Biran et al., 2009). Similarly, Mercy Corps partnered with SweetSense to use motion sensors with water flow sensors to monitor the prevalence of handwashing after latrine use. They found that water use after latrine use was very low (¡ 10%) in all but one district, which registered almost 40% use of water after latrine use. They also found that self-reported use of the latrine and handwashing after using the latrine were much greater (up to 4 times and 25 times, respectively) than the latrine use and handwashing after latrine use detected by the sensors (Thomas and Mattson, 2013).

Finally, video observation can be an effective tool for observing and recording handwashing behaviour unobtrusively. Although no comparisons have been conducted, it is possible that video observation would be preferable to direct observation in settings where handwashing behaviour can be clearly recorded from a fixed location. Video observation has the advantage of being able to record over longer time periods without interruption, and recordings can be reviewed rapidly by a human observer. It is also possible that discreetly placed video observation may reduce reactivity, although there are ethical concerns that must be considered when consent cannot be obtained for all involved parties. Like sensors, video observation provides a dynamic measure of handwashing behaviour over time, but it also allows the reviewer to differentiate the handwashing behaviours of specific individuals (Pickering et al., 2014).
2.6.4 Equitable / Acceptable

Like sanitation, the goals of equity and acceptability in hygiene practices may require more qualitative methodologies like ethnography and semi-structured interviews to understand what motivates hygienic behaviour, to gauge the acceptability of hygiene interventions, and to characterize intra-household hygiene behaviours. This is particularly true for the special needs of women and girls and the ambiguity surrounding indicators for menstrual hygiene management. For example, Curtis et al. found that social affiliation and disgust were two strong motivators of handwashing behaviour, but that fear of disease had little influence on handwashing behaviour (Curtis et al., 2009).

As proposed by Ram, composite measures would ideally be employed to more accurately characterize handwashing behaviours (Ram, 2013). While some methodologies are particularly suited to measuring specific outcomes (e.g. sensors for monitoring handwashing practices), no one methodology is adequate for verifying and monitoring all four hygiene outcomes.

2.7 Combined Methodologies

Given that all monitoring and evaluation methods have their own advantages and limitations, it is often beneficial to use more than one method to get a fuller picture of WASH behaviour. Combined methodologies can reinforce the advantages while mitigating the limitations of each of the monitoring techniques. Surveys, ethnographies and direct observation give context to sensor readings that remain objective despite interim analyses. Using more than one monitoring method also increases the level of detail available. Sensors or spot checks may give a picture of household characteristics, but surveys and structured observations can be used to inform individual
behaviour. In turn, these combined sources provide further refinement for sensor algorithms or survey indices for streamlined analysis during subsequent monitoring periods. Comparisons and correlations across different monitoring methodologies can also be used to support the internal validity of results.

An important function of sensors is their ability to validate the reliability of another method while also suggesting improvements to standards of practice. For example, higher correlation between sensors and self-reports is seen when the questionnaire focuses on near-time behaviour, particularly in the previous 48 hours, suggesting how surveys should be administered in future studies (Sinha et al., 2016). The appropriateness of structured observation as the gold standard is also questionable given reactivity that has been observed with passive latrine use monitors (Clasen et al., 2012). Similar findings were made with SmartSoap for handwashing behaviour (Ram et al., 2010). Integrated methods over appropriately sampled sub-populations could greatly reduce measurement and inferential uncertainty. For example, DelAgua Health Ltd’s monitoring and evaluation programme linked household surveys and indicators, health outcomes and sensors to determine the impact of a cookstove and water filter intervention in Rwanda (Thomas and Mattson, 2013; Rosa et al., 2014; Barstow et al., 2016; Nagel et al., 2016; Snoad et al., 2017). Although stronger monitoring methodologies and technologies can be cost-prohibitive at scale, continued cost reductions, widespread network coverage for mobile devices, and increased access to and familiarity with the Internet of Things and mobile devices could facilitate increased use of information and communication technologies for monitoring. When the budget, time and training are available, combined methodologies can provide a more comprehensive and instructive depiction of WASH usage.
2.8 Conclusion

Although the SDG target for sanitation and hygiene represents a marked improvement over the MDG target, there is still a substantial disconnection between the desired improvements in service levels and the proposed indicators. The inclusion of safety, adequacy and equity acknowledge that progress cannot be measured by simply counting the number of latrines or soap bars. Instead, health benefits from improved sanitation and hygiene facilities depend on the facilities’ accessibility, their use, and their safety – their ability to effectively separate excreta from human contact both within and beyond the household. However, the proposed service ladders still rely heavily on direct observation of sanitation and handwashing facilities to infer usage, the safety of the facility, and the management of excreta. Similarly, household surveys depend on self-reported estimates of use, acceptability and equity that can differ significantly from actual use due to courtesy or recall bias. This combination of measurement error based on self-reporting and inferential error based on facility type can compound the uncertainty associated with estimates of real health impact due to improved sanitation and hygiene facilities. Technologies or methodologies that accurately measure sanitation outcomes—i.e. use, household safety, community safety, etc.—can reduce uncertainty by reducing measurement error and limiting the chain of causal inference between the indicator and the desired effect.
References


the performance of indicators of hand-washing practices in rural Indian households.


A. J. Pickering, A. G. Blum, R. F. Breiman, P. K. Ram, and J. Davis. Video surveillance captures student hand hygiene behavior, reactivity to observation, and peer


60


Chapter 3

Toilet Alarms: A Novel Application of Latrine Sensors and Machine Learning for Optimizing Sanitation Services in Informal Settlements

Abstract

The cost-effectiveness and reliability of waste collection services in informal settlements can be difficult to optimize given the geospatial and temporal variability of latrine use. Daily servicing to avoid overflow events is inefficient, but dynamic scheduling of latrine servicing could reduce costs by providing just-in-time servicing for latrines. This study used cellular-connected motion sensors and machine learning to dynamically predict when daily latrine servicing could be skipped with a low risk of overflow. Sensors monitored daily latrine activity, and enumerators collected solid and liquid waste weight data. Given the complex relationship between latrine use and the need for servicing, an ensemble machine learning algorithm (Super Learner) was used to estimate waste weights and predict overflow events to facilitate dynamic scheduling. Accuracy of waste weight predictions based on sensor and historical weight data was adequate for estimating latrine fill levels (mean error of 20% and 22% for solid and liquid wastes), but there was greater accuracy in predicting overflow events (area under the receiver operating characteristic curve of 0.90). Although our simulations indicate that dynamic scheduling could substantially reduce costs for lower use latrines, we found that cost reduction was more modest for higher use latrines and that there was a significant gap between the simulated and implemented results.

Keywords: sanitation, passive latrine use monitors (PLUMs), machine learning, information and communication technologies (ICTs), Super Learner
3.1 Introduction

Globally, at least 2.3 billion people do not have access to improved sanitation facilities, and 4.5 billion people do not have access to safely managed sanitation services (UNICEF / WHO, 2017). While much attention has been focused on latrines for rural populations and campaigns to end open defecation (UNICEF / WHO, 2017; Robiarto et al., 2014; Trémolet, 2011; Coffey et al., 2014), the need for improved and safely managed sanitation facilities is acute in dense informal settlements in rapidly urbanizing areas (Bohnert et al., 2016; Brown et al., 2015). This need has three principal drivers: the high population density of informal settlements, the lack of institutional sanitation providers, and the challenge of safely transporting fecal waste out of the settlement (Paterson et al., 2007; Mara, 2012).

Today, more than half of humanity lives in a city. In low income countries the trend toward urban migration is particularly strong, with 31% of the population residing in urban areas and 4.2% of the population migrating to cities each year (United Nations Department of Economic and Social Affairs, 2015). However, urban growth and infrastructure development has often not been able to keep pace with the rapid influx of individuals and families, resulting in the formation of informal settlements and squatter’s communities that lack basic water, sanitation, or electrical services (United Nations, 2015). The lack of sanitation services in informal settlements is particularly
problematic, as fecal deposition in high traffic environments combined with increased residential density can greatly increase the risk of enteric infections (Kimani-Murage et al., 2014; Bhagwan et al., 2008). For example, children in Nairobi’s informal settlements have a prevalence of diarrhea (20.2%) that is comparable to prevalences in rural Kenya (21.7%) but much greater than the rate reported for Nairobi at large (14.8%) (African Population and Health Research Center, 2014).

Attempts to provide reliable and appropriate sanitation services in informal settlements are often limited by the lack of legal protections, property ownership, resistance from governing authorities, and minimal water and sewage infrastructure (Bohnert et al., 2016). Given the lack of support from governments, sanitation solutions in informal settlements often depend on non-profits or social enterprises that rely on donations or revenue generating models to sustain services (Auerbach, 2016).

One of the key factors influencing the cost-effectiveness and reliability of service provision in informal settlements is the ability to optimize waste collection from latrines with variable use patterns that are spatially dispersed within an informal settlement. Optimization of latrine servicing typically implies a trade-off between increased collection efficiency and increased risk of latrine overflow events. Daily servicing effectively avoids the risk of latrine overflow, but inefficient servicing of latrines (i.e., servicing latrines before they are full) may not be cost-effective. On the other hand, less frequent servicing increases the likelihood of a latrine overflow event, which can be damaging to the operator’s reputation, result in decreased demand or willingness-to-pay for services, as well as increase the risk of exposure to fecal contamination. Ideally, latrines would be serviced with the highest efficiency possible, but to do so requires real- or near-time monitoring of latrine fill levels (i.e., the fullness of the solid and liquid waste receptacles). In previous studies motion detector sensors (passive latrine use monitors - PLUMs) have been used to monitor latrine activity.

66
and compared against self-reported latrine use or observed latrine use (Delea et al., 2017; Bohnert et al., 2016; Sinha et al., 2016; O’Reilly et al., 2015). However, there are no known studies that attempt to estimate the accumulated solid or liquid waste detected using a latrine sensor.

Partnering with Sanergy Inc., an established sanitation service provider for informal settlements in Nairobi, Kenya, researchers from Portland State University and Sweet Sense investigated how latrine sensors could be used to estimate waste fill levels and improve servicing efficiency for forty latrines in Nairobi, Kenya. In particular, we evaluated (1) how accurately we could estimate solid and liquid waste weights based on motion sensor data, (2) how accurately we could predict a latrine overflow event to create a dynamic schedule for latrine servicing, and (3) how cost-effective sensor-enabled servicing would be compared to daily servicing or servicing based on data from on-site weighing. In order to answer these questions we developed three models to simulate the predictive performance and cost-effectiveness of dynamic scheduling in relation to Sanergy’s existing static schedule. We also present the results from a dynamic schedule that was implemented over three months and compare its performance to the existing and simulated scheduling scenarios.

3.2 Materials and Methods

For this study a convenience sample of forty latrines was selected for installing the motion sensors. These forty latrines were chosen because they were clustered along a service route that was close to the central office and had reliable waste collector personnel. Forty-one latrines from a nearby route were selected as the comparison group to estimate outcome variables at baseline and after the intervention (see Table 3.1). General characteristics of each latrine were obtained from Sanergy’s existing records (i.e., type of latrine, responsible waste collectors and field officers, and collection
In addition, three enumerators were employed to manually weigh and record daily on-site solid and liquid waste weights each time a latrine was serviced in the intervention and comparison groups. Weight measurements were recorded using the following procedure: (1) enumerators accompanied waste collectors each morning to each of the latrines designated for servicing; (2) at each latrine waste collectors removed the solid and liquid waste cartridges and weighed each cartridge using a hanging scale (see TOC image); (3) weights were manually recorded by the enumerators using a mobile application that did not rely on cellular network connectivity; (4) weight measurements were uploaded to the survey server each afternoon when enumerators returned to the main office; (5) an automated algorithm compiled weight records from the survey, subtracted the weight of the empty solid and liquid waste cartridges, and compared the list of latrines serviced against the list of latrines scheduled for servicing to account for missing data or discrepancies. Enumerators were also responsible for installing, trouble-shooting, and swapping out sensors when batteries were running low or sensors were not reporting. Sensors were installed in October, 2016, and three months of baseline weight and sensor data were collected before the intervention period from January through March, 2017. During the baseline period, all latrines were scheduled for servicing according to Sanergy’s static schedule, whereas during the intervention period latrines with sensors were serviced using a dynamic schedule (both schedules described in further detail below). The purpose of the experiment was to see whether collection efficiency improved in the latrines with sensors during the intervention period when weight and sensor data were used to generate a dynamic servicing schedule.

The sensor unit was equipped with a passive infrared motion sensor that logged movement in the latrine throughout the day and transmitted the data each evening
via a GSM radio to Sweet Sense servers. After all the sensors had called in, an automated algorithm was executed to compile all the weight and motion sensor data and run the machine learning algorithm to determine which latrines could be skipped the next day. During the intervention period, waste collectors were notified via text message each morning which latrines to skip. The sensor unit was also equipped with an RFID reader that logged activity from the waste collectors. Waste collectors were
instructed to swipe their “Collected” or “Not Able to Collect” tags depending on the action taken. The “Not Able to Collect” tag was reserved for instances when the facility had overflowed or required cleaning beyond the waste collector’s responsibility, but there were no instances when the “Not Able to Collect” tag was used. The latrine operator was also given an RFID tag to request assistance, and RFID scans from latrine operators were immediately transmitted to Sweet Sense servers and triggered a Salesforce push notification for Sanergy staff to check-in with the latrine operator. Finally, sensor data were uploaded to the Sweet Sense dashboard to display the daily collection schedule, the log of Salesforce push notifications and waste collector scans, and the approximate number of uses for each latrine.

In order to measure changes in the efficiency of latrine servicing over the course of the intervention period, the average solid waste fill level and capacity savings were selected as the main outcome variables. Waste fill level as a percent was defined as follows:

\[
\text{Fill Level} = \frac{\text{Waste Weight}}{\text{Waste Density}} \times \frac{1}{\text{Cartridge Capacity}}
\]  

(3.1)

<table>
<thead>
<tr>
<th>Table 3.1: Sample Characteristics</th>
<th>sensor</th>
<th>no sensor</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of latrines</td>
<td>40</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>number of observations</td>
<td>4870</td>
<td>4797</td>
<td></td>
</tr>
<tr>
<td>collections per latrine: median (IQR)</td>
<td>141 (32)</td>
<td>133 (21)</td>
<td>0.331</td>
</tr>
<tr>
<td>solid waste container sizes</td>
<td>31 with 45L</td>
<td>41 with 40 L</td>
<td></td>
</tr>
<tr>
<td>high use latrines: number (%)</td>
<td>21 (52%)</td>
<td>11 (27%)</td>
<td></td>
</tr>
<tr>
<td>low use latrines: number (%)</td>
<td>19 (47%)</td>
<td>30 (73%)</td>
<td></td>
</tr>
<tr>
<td>solid waste fill level: median (IQR)</td>
<td>0.52 (0.23)</td>
<td>0.43 (0.24)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>liquid waste fill level: median (IQR)</td>
<td>0.41 (0.20)</td>
<td>0.34 (0.20)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Waste weights were determined by weighing solid and liquid waste cartridges on-site at the time of servicing, and the cartridge weight was subtracted from the waste weight using an automated algorithm. While the density of the solid waste varied based on the amount of sawdust and toilet paper used, a conservative density of 0.721 kilograms per liter was used to convert solid waste weight to solid waste volume based on the average weight recorded for full cartridges. The solid waste volume was then divided by the cartridge capacity, which varied between 40 L and 45 L, to determine the latrine fill level (see Equation 3.1). Given that solid waste generally filled faster than liquid waste, the average solid waste fill level was selected as the primary outcome variable for measuring changes in servicing efficiency. Capacity savings were defined as the number of latrine servicing events that could be avoided due to dynamic scheduling.

3.2.1 Predictive Models

We initially assumed that estimates of latrine fill levels based on motion sensor data would be sufficient for predicting when latrines could be skipped. However, while we were able to predict waste fill levels with sufficient accuracy (mean absolute percent error of 20% and 22% for solid waste and liquid waste, respectively), we found that the motion sensor data on their own (i.e., the number of recorded movements or clicks from the motion sensor) were not sufficient to accurately predict whether the latrine needed to be serviced the next day. Figure 3.2 attempts to characterize the complex chain of factors that make latrine servicing predictions difficult. First, waste weights did not always accurately reflect waste volumes because of the variable amount of consumables that were used each day (i.e., the amount of sawdust and toilet paper present in the solid waste cartridge) and the different cartridge volumes in each latrine. Second, the need to be serviced depended not only on the estimated
fill level from the first day’s latrine activity, but also on the anticipated waste that would be added the next day if the latrine were skipped. Also, conversations with latrine operators revealed that full cartridge capacity was not always desirable due to increased odor and complaints from customers. Finally, even when it was determined that a latrine needed to be serviced, there was no guarantee that the waste collector would service the latrine. Sometimes waste collectors were not able to access latrines, and sometimes waste collectors used their own judgment based on a visual inspection of the fill level and their experience with the route to determine whether the latrine needed servicing. Waste collectors also indicated that they were more likely to service some latrines based on the preferences of the operator, often creating a tension between Sanergy’s desire for more efficient servicing and the operators’ desires for more frequent servicing.

Figure 3.2: Chain of factors contributing to a latrine’s need to be serviced.

Given the complex relationship between latrine use and servicing demand, we established that a simple linear correlation between motion sensor data and estimated fill levels would be insufficient for accurately predicting the need for servicing. Instead we used a machine learning algorithm (Super Learner, Polley et al., 2016) to predict when latrines would need to be serviced based on a variety of features that were identified using the available data (see Figure 3.3). We developed four models to
compare the accuracy and cost-effectiveness of different scheduling scenarios. The first model represented Sanergy’s business-as-usual static schedule, and the three simulated models represented the performance of dynamic scheduling using different data sources. In addition, we present in Table 3.2 the results from the actual dynamic schedule that was used during the intervention period and an additional simulated scenario that applies dynamic scheduling to lower-use latrines.

For the first model (Static Schedule) we used Sanergy’s existing servicing schedule where thirty-six latrines were serviced daily and four latrines had reduced servicing schedules (i.e., four latrines were only serviced on Sundays, Mondays, Wednesdays, and Fridays based on waste collector recommendations). A dichotomous outcome variable was created to model whether a latrine would have overflowed had it been skipped based on weight data from consecutive days (i.e., if the estimated volumes from two consecutive days exceeded the cartridge capacity, then the outcome variable was classified as one; otherwise it was classified as zero).

In the second model (Sensor Only), we used sensor data and the Super Learner algorithm to predict when latrine servicing could be skipped. The predictor variables for this model included the latrine ID, the day of the week, and the normalized number of clicks from the motion sensor in the latrine. In addition, we used the number of clicks to create features that approximated the number of latrine uses and the number of edges associated with latrine use based on the methodology described in Clasen et al. (2012). This scenario was used to simulate the performance and cost-effectiveness of dynamic scheduling without the daily enumeration of weight data and servicing events.

For the third model (Weight Only), we used the record of daily solid and liquid waste measurements to predict when latrine servicing could be skipped. We first used Super Learner to predict the solid and liquid waste weights based on historical weight
data (i.e., the latrine ID, the day of the week, and previous weight data collected from that latrine). Given the variability of latrine fill levels throughout the week, we created several features that improved the model’s performance in predicting latrine waste weights, including: the average weight for each day of the week, the average weight for the previous seven days, the average weight for the previous three days, the weight from the previous day, and the first quartile, third quartile, median, and average overall weights for each latrine. The weight predictions from the first layer of the algorithm were then incorporated as a feature in the second layer of the algorithm that was used to predict the probability of an overflow event if skipped. This scenario was used to simulate the performance of dynamic scheduling with on-site weighing but without the capital and operating expenses associated with the sensors.

Finally, the fourth model (Sensor+Weight) combined sensor and weight data to predict waste weights and then used the full set of features to predict the need for servicing. Predictions from the fourth model were used for dynamic scheduling during the implementation period, and we describe below the additional safeguards that were incorporated to prevent overflows. Finally, the relative importance of each of the features used in the three prediction models is shown in Figure 3.3.

### 3.2.2 Evaluation of Prediction Models

All four models were evaluated using R (R Development Core Team, 2011), including the ROCR (Sing et al., 2009) and SuperLearner (Polley et al., 2016) packages. Super Learner is an ensemble learner that employs a variety of screening and prediction algorithms to improve the accuracy of prediction (Polley and van der Laan, 2010). It has been used in recent studies to predict the failure of rural handpumps (Wilson et al., 2017) as well as to predict virological failure for HIV-positive patients on antiretroviral therapy (Petersen et al., 2015).
Figure 3.3: Relative importance of features used in the learner for predicting the probability of an overflow event for solid waste. The relative importance represented above is based on the mean decrease in Gini impurity from the randomForest learner. Gini impurity refers to the improvements in data classification that are contributed by each feature (Archer and Kimes, 2008).

Several learners used to predict continuous and binomial outcomes were incorporated, including (ordered by weighting): Lasso regression (Tibshirani, 1996), multivariate adaptive regression splines (Hastie and Tibshirani, 1987; Milborrow, 2018), and random forests (Friedman, 2001). In order to evaluate the performance of each prediction model, the data were randomly split into training and testing sets based on each latrine site (70:30). To determine the relative weights associated with each learner’s prediction in the ensemble, the algorithm performed ten-fold cross validation using the training data. The algorithm’s predictive performance was then evaluated using the test data, where the mean absolute percent error (MAPE) was used to evaluate continuous outcomes and the area under the receiver operating characteristic (AUROC) curve, accuracy, sensitivity, and specificity were used to evaluate classification performance. The AUROC was selected as the primary metric for model
comparison because it captures the overall accuracy of the model in predicting outcomes, regardless of the threshold chosen (see below), where an AUROC equal to one indicates perfect classification.

For the purpose of this investigation the number of true negatives (i.e., instances when the algorithm accurately predicted that a latrine would not overflow if service were skipped) represented the potential for cost-savings due to higher efficiency latrine servicing. Given that the algorithm output a probability of overflow ranging from zero to one, a threshold was selected that would provide the lowest number of false negatives (i.e., instances when the algorithm incorrectly predicted that a latrine could be skipped) while minimizing the number of false positives (i.e., instances when the algorithm incorrectly predicted that a latrine had to be serviced). We were unable to quantify the overall cost of a false negative or latrine overflow event, as it involved tangible costs (e.g., latrine servicing crew, cleaning supplies, lost revenue due to latrine being closed, etc.) as well as intangible costs (e.g., damage to reputation of Sanergy brand or latrine operator, exposure to fecal contamination, etc.). As a result, we chose a threshold that allowed for the fewest number of potential overflow events, where potential overflow events were defined as latrine fill levels that were between 1.00 and 1.10 capacity.

3.2.3 Cost Assumptions

Servicing costs for each scenario were estimated based on cost and logistics data provided by Sanergy. Given that the primary expense for latrine servicing is labor, and given the small sample size for this experiment, costs were simplified to a per servicing event estimate. Cost-savings are represented as the amount of time and labor that could be avoided if dynamic scheduling were adopted at scale for latrines with similar use patterns.
3.3 Results

Over the course of six months 4,870 service events were recorded for the forty latrines with sensors. When merged with the sensor data, a total of 4,371 weight and sensor observations were available for training and testing the learner. As seen in Figure 3.4 and Table 3.2, overall classification performance of the Static Schedule was low (AUROC of 0.52), whereas classification performance increased dramatically with the additional information provided by sensors (0.87), historical weight data (0.89), and combined sensor and weight data (0.90). Figure 3.5 displays the sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) that were evaluated on the testing data that was not used in model fitting. In addition, Table 3.2 displays the simulated performance of each model during the intervention period from January through March, 2017, including the predicted number of skips, the number of possible overflows, the capacity savings due to decreased latrine servicing, and the estimated savings per month based on reduced costs for labor and consumables. In total, there were 2,272 servicing events recorded during the three-month intervention period for the latrines with sensors. There were 566 opportunities for skipping servicing, and the performance of each of these models in predicting these potential skips varied considerably. Sanergy’s static schedule reflected approximately 2% of the possible skips, whereas the dynamic schedules using sensor and weight data were able to predict between and 12% and 13% of the possible skips.
Table 3.2: Performance metrics for the four prediction models, the actual implementation results, and a prediction model using low-use latrines.

<table>
<thead>
<tr>
<th>Model Performance</th>
<th>Static Schedule</th>
<th>Sensor Only</th>
<th>Weight Only</th>
<th>Sensor+Weight</th>
<th>Actual Schedule&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Low-Use Latrines&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance on Test Data From Baseline and Intervention Periods</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sensitivity</td>
<td>100%</td>
<td>96.4%</td>
<td>97.3%</td>
<td>97.9%</td>
<td>99.2%</td>
<td>95.4%</td>
</tr>
<tr>
<td>specificity</td>
<td>4.50%</td>
<td>53.7%</td>
<td>61.2%</td>
<td>61.9%</td>
<td>6.23%</td>
<td>63.1%</td>
</tr>
<tr>
<td>positive predictive value</td>
<td>49.2%</td>
<td>65.9%</td>
<td>69.9%</td>
<td>70.5%</td>
<td>55.5%</td>
<td>50.3%</td>
</tr>
<tr>
<td>negative predictive value</td>
<td>100%</td>
<td>94.2%</td>
<td>96.0%</td>
<td>97.0%</td>
<td>86.7%</td>
<td>97.2%</td>
</tr>
<tr>
<td>accuracy (AUROC)</td>
<td>52.2%</td>
<td>86.6%</td>
<td>89.2%</td>
<td>89.5%</td>
<td>52.7%</td>
<td>90.5%</td>
</tr>
<tr>
<td>Performance on All Data During Three-Month Intervention Period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>predicted skips</td>
<td>46&lt;sup&gt;c&lt;/sup&gt;</td>
<td>279&lt;sup&gt;c&lt;/sup&gt;</td>
<td>274&lt;sup&gt;c&lt;/sup&gt;</td>
<td>298&lt;sup&gt;c&lt;/sup&gt;</td>
<td>75&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1142&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>possible overflow events</td>
<td>0</td>
<td>47</td>
<td>17</td>
<td>18</td>
<td>10&lt;sup&gt;f&lt;/sup&gt;</td>
<td>69</td>
</tr>
<tr>
<td>capacity savings&lt;sup&gt;g&lt;/sup&gt;</td>
<td>2.0%</td>
<td>12%</td>
<td>13%</td>
<td>13%</td>
<td>3.3%</td>
<td>52%</td>
</tr>
<tr>
<td>waste collector labor&lt;sup&gt;h&lt;/sup&gt;</td>
<td>$1100</td>
<td>$1000</td>
<td>$1000</td>
<td>$990</td>
<td>$1100</td>
<td>$530</td>
</tr>
<tr>
<td>total consumables&lt;sup&gt;i&lt;/sup&gt;</td>
<td>$150</td>
<td>$140</td>
<td>$140</td>
<td>$140</td>
<td>$150</td>
<td>$73</td>
</tr>
<tr>
<td>total cost per quarter</td>
<td>$1300</td>
<td>$1100</td>
<td>$1100</td>
<td>$1100</td>
<td>$1300</td>
<td>$600</td>
</tr>
<tr>
<td>savings per month&lt;sup&gt;j&lt;/sup&gt;</td>
<td>NA</td>
<td>$44</td>
<td>$43</td>
<td>$48</td>
<td>$5</td>
<td>$200</td>
</tr>
</tbody>
</table>

<sup>a</sup> Performance for Actual Schedule is based on the dynamic schedule from the implementation period.

<sup>b</sup> Performance of the weight only model on lower use latrines in the comparison group.

<sup>c</sup> Out of 566 possible skips.

<sup>d</sup> Represents the actual number of skips during the intervention period.

<sup>e</sup> Out of 1383 possible skips.

<sup>f</sup> Instances when a latrine was scheduled for a skip but waste collectors serviced the latrine based on visual inspection of fill-level; there were no reported overflow events during the baseline or intervention periods.

<sup>g</sup> Number skips divided by the total number of servicing days.

<sup>h</sup> USD per quarter based on Sanergy records, with the average waste collector servicing 15 latrines per day and receiving a monthly salary of USD $225.

<sup>i</sup> USD per quarter based on USD $0.08 for disposable bags, sanitary bags, water, cleaning, and incineration per service event.

<sup>j</sup> Saving compared to the static schedule.
3.3.1 Comparison Group

Over six months 4,797 service events were recorded for the forty-one latrines without sensors that served as a comparison group. As shown in Table 3.1, the latrines with sensors had a higher median fill level compared to the latrines without sensors (52% vs. 43%). Given that the majority of the latrines with sensors were high-use latrines, where high-use was defined as having a maximum fill level and a third-quartile fill level greater than 60% of the cartridge capacity, there was less room for improving efficiency in the latrines with sensors compared to the comparison group. That is, the fact that latrines were generally 52% full meant that there were fewer opportunities for skipping the latrines with sensors compared to the latrines without sensors. Despite there only being a 9% difference in median fill levels between the two groups there was significantly more opportunity for skipping in the comparison group. Using only weight data, the Super Learner algorithm was able to predict 1,142 skip events with a
Figure 3.5: Sensitivity (Sens), specificity (Spec), negative predictive value (NPV), and positive predictive value (PPV) for solid waste overflow predictions over a range of probability thresholds.

A high degree of accuracy (AUROC of 0.91) and an estimated capacity savings of 52%. Given that we were not able to test dynamic scheduling in the comparison group, these simulated results represent the upper bound of potential capacity savings. As seen in Figure 3.6, average fill levels for latrines in both groups increased over the intervention period, which may reflect seasonal trends or general uplift due to Sanergy’s efforts to improve servicing efficiency over the same period.

3.4 Discussion

Using weight and sensor data from forty latrines in an informal settlement in Nairobi, we were able to demonstrate that a machine learning algorithm can predict with a high degree of accuracy when latrine servicing could be skipped (AUROC from 0.87 to 0.90 and capacity savings from 12% to 13%). These predictions were then used
Figure 3.6: Average fill levels for the latrines with sensors (dashed line) and the latrines without sensors (solid line) for the baseline (pink) and intervention (blue) periods.

to create a dynamic latrine schedule that increased solid waste collection efficiency by 5% between the baseline and intervention periods (see Figure 3.6). Although the machine learning algorithm was more effective in identifying skip events compared to the Static Schedule (AUROC 0.52 and capacity savings of 2%), there was a significant gap between the simulated performance of the algorithm and the implemented results (AUROC 0.53 and capacity savings of 3%). We attribute this gap to implementation challenges as well as the diminished opportunity for skipping associated with higher use latrines.

Implementation challenges were numerous. Dynamic scheduling represented a significant deviation from the static schedules that waste collectors and field staff were accustomed to. Collecting accurate weight data was difficult, and the standard practice of recording cartridge weights at a central weighing station was prone to error. A set of two on-site weighing machines were fabricated to facilitate weight
measurements at the time of servicing, but data entry was still subject to human error (e.g., inaccurate designations of latrines, entry error, delayed uploading of records to the server, etc.). In addition, there were initially no records that were logged for latrines that were skipped, so it was impossible to distinguish between latrines that were skipped and data that were missing. This was corrected by creating a new mobile survey for waste records and an automated algorithm to check that events were logged for each latrine. However, even with these redundancy measures about 5% of expected entries were not accounted for each day.

Because the dynamic schedule was new and required the approval and cooperation of latrine operators, the algorithm was initially tuned conservatively in order to minimize the risk of an overflow event. For example, even though solid wastes were the primary driver of service events, a probability of overflow for either solid or liquid wastes automatically designated a latrine for collection. In addition, if a latrine was skipped or there was a missed entry from the previous day, the latrine was automatically scheduled for collection. However, we eventually realized that waste collectors often skipped low-use latrines regardless of scheduling. Since missing data entries automatically designated a latrine for collection, lower-use latrines were often scheduled for collection even when waste collectors knew that they could be skipped. This combination of missing data and conservative scheduling resulted in a general distrust in the algorithm’s predictions, prompting many waste collectors to service latrines according to their own intuition rather than the dynamic schedule.

However, it is important to note that the waste collector’s intuition was correct more often than not. On at least ten occasions, the algorithm scheduled a latrine for skipping that clearly would have overflowed had the waste collector not serviced the latrine based on visual inspection. In this regard, the route selected for installing sensors was a safe choice because the waste collectors were reliable and the route was
well-known and accessible by Sanergy staff. However, these very attributes also made the route less useful for the experiment, as the information being provided by the sensors and daily weights was unnecessary given the familiarity of the waste collectors and the daily servicing needed by most latrines. As a result, it was determined that collecting data from sensors or daily weights would be most useful on new routes where latrine patterns were still being established, on existing routes where latrine use was more variable, or on routes where latrines were used less frequently.

Although the accuracy of the algorithm may not be much better than that of a seasoned waste collector, there is an additional advantage that motion sensor data, weight data, or RFID scans can provide: the ability to track latrine servicing. Sanergy’s capacity for reallocating waste collector labor depends on its ability to predict when latrines will need to be serviced while reliably tracking when latrines have been serviced. In this way service records provide a form of accountability for waste collectors, a quality assurance mechanism for honoring contracts with latrine operators, and a dataset for predicting future servicing. However, the high cost of hardware relative to the low cost of labor in Nairobi implies that cost savings would need to significantly increase for Sanergy to implement any changes at scale. Our simulations suggest that sensor and weight measurements could save between $43 and $200 per month for a route with approximately forty latrines depending on the frequency of use of the latrines. This cost savings represents the upper bound on all expenses related to latrine sensors (e.g., hardware, data transmission, operation and maintenance personnel, predictive analytics), weight records (e.g., enumerators, mobile devices, and predictive analytics), or RFID scanners. However, given the gap between simulation and implementation, these estimates may be optimistic.

There are additional considerations that may temper the cost savings associated with dynamic scheduling. First, 92% of the latrines with sensors and 54% of the
latrines without sensors were co-located, meaning that latrines were being managed by the same operator in clusters of two or three. Co-located latrines were more likely to be skipped compared to standalone latrines, but the benefit of skipping a latrine is greatly diminished if waste collectors are already servicing a latrine in the same location. Second, this analysis was not able to quantify the potential cost associated with an overflow event. This cost would include additional labor and supplies for servicing an unsanitary latrine, but it would also include damage to the operator or Sanergy’s reputation and reduced patronage. In addition, the current algorithm uses the latrine ID as a predictor variable to capture site-level variability and latrine-use trends. However, using the latrine ID as a predictor also makes the algorithm less portable given the need to collect baseline data from new latrines before making predictions on a new route. However, this baseline burn-in may be inevitable given that average weight trends were also significant predictors in the algorithm. Finally, this analysis was not able to take into consideration the additional administrative cost associated with reallocating waste collectors in a dynamic scheduling scenario. Given the geospatial distribution of latrines, the inability to remotely chart pathways through informal settlements, and challenges finding and accessing latrines for waste collection, it would be exceedingly difficult to dynamically redraw servicing routes for waste collectors on a regular basis.

In this study, sensors were able to monitor latrine activity, track latrine servicing, and facilitate communication between Sanergy staff and latrine operators. However, sensor data did not significantly improve the algorithm’s performance compared to weight data alone. Regardless, this study provides a promising application of machine learning for estimating waste weights and dynamically scheduling latrine servicing. Although we found that implementation lagged simulation significantly, we anticipate a much greater potential for servicing efficiency and cost savings when applied to lower
use latrines.

Acknowledgements

We appreciate our partnership with Sanergy Inc., and in particular the enumerators, waste collectors, field staff, and administrators that made this study possible. We also want to thank Jeremy Coyle for reviewing the code used in this analysis. This study was supported by the United Kingdom Department for International Development through the GSM Association, the Link Foundation, and the National Science Foundation IGERT Grant #0966376: “Sustaining Ecosystem Services to Support Rapidly Urbanizing Areas.” Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
References


urban advantage has seemingly been wiped out. *Health and Place*, 29:95–103, 2014. ISSN 18732054. doi: 10.1016/j.healthplace.2014.06.003. URL http://dx.doi.org/10.1016/j.healthplace.2014.06.003.


E. Polley, E. LeDell, and M. van der Laan. Package ‘SuperLearner’: Super


Chapter 4

Improved Drought Resilience Through Continuous Water Service Monitoring and Specialized Institutions - A Longitudinal Analysis of Water Service Delivery Across Motorized Boreholes in Northern Kenya

This chapter has been accepted for publication as an article in the journal Sustainability.

Abstract

Increasing frequency and severity of drought is driving increased use of groundwater resources in arid regions of Northern Kenya, where approximately 2.5 million people depend on groundwater for personal use, livestock, and limited irrigation. As part of a broader effort to provide more sustainable water, sanitation, and hygiene services in the region, we have collected data related to site functionality and utilization for approximately 120 motorized boreholes across five counties. Using a multilevel model to account for geospatial and temporal clustering, we found that borehole sites that counties had identified as strategic assets during drought, ran on average about 1.31 hours less per day compared to non-strategic borehole sites. As this finding was contrary to our hypothesis that strategic boreholes would exhibit greater utilization on average compared to non-strategic boreholes, we consider possible explanations for this discrepancy. We also use a coupled human and natural systems framework to explore how policies and program activities in a complex system depend on consistent and reliable feedback mechanisms. Funding provided by the United States Agency for International Development. The views expressed in this article do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

Keywords: water services, remote monitoring, sensors, coupled human and natural systems, Kenya
4.1 Introduction

Approximately 3.6 billion people live in regions that are water-scarce at least one month per year, a condition expected to increase to between 4.8 and 5.7 billion people by 2050 (WWAP, 2017). In part, water insecurity is attributable to the effects of climate change that are already altering seasonal precipitation cycles around the globe. Whereas rainfall is increasing in higher latitudes, land areas south of the Sahara in Africa have seen records of dry months increasing by as much as 50%, despite predictions of increased precipitation in Eastern Africa based on climate models (Lehmann et al., 2018; Rowell et al., 2015). Meanwhile, warmer temperatures are drying soils, leading to a reduction in the amount of precipitation being absorbed in surface water sources, catchments, and aquifers (Shah and Mills, 2018). The combination of greater precipitation variability, longer dry seasons, and reduced water storage represent a substantial natural hazard for the 2.5 million individuals residing in the arid and semi-arid regions of Kenya that increasingly depend on ground water resources for personal consumption, livestock, and limited irrigation during the dry season.

However, there is an important distinction between the natural hazard of drought as a hydrological imbalance (Banholzer et al., 2014) and the risk of drought disaster that is a function of a region’s vulnerability and capacity to cope (Fitzgibbon and Crosskey, 2013). Drought emergencies are characterized by a confluence of issues including: lack of access to water, lack of access to pasture for feeding livestock, competition among communities for scarce resources, and insecurity because of conflict over resources (NDMA, 2015). The cost in human lives and humanitarian aid has been significant for each of the drought emergencies experienced in Eastern Africa. For example, the five droughts in Kenya between 1998 and 2011 resulted in approxi-
mately 1.6 billion USD in humanitarian aid, and each drought affected between two and four million people (Fitzgibbon and Crosskey, 2013).

This study presents the initial findings from two initiatives aimed at decreasing vulnerability to drought emergencies in five Northern Kenyan counties. We first provide a brief description of both initiatives, followed by a presentation of the theoretical framework that informs this investigation. We then present the results of a multilevel analysis that explores the factors influencing water system utilization in these five counties. Finally, we conclude with a discussion of the implications of this investigation and the opportunities for further research.

4.1.1 Kenya RAPID Program

The Kenya Resilient Arid Lands Partnership for Integrated Development (Kenya RAPID) program is aimed at increasing “access to water and sanitation for people and water for livestock” while rebuilding “a healthy rangeland-management ecosystem” (USAID, 2018). Kenya RAPID is funded by the United States Agency for International Development (USAID) and the Swiss Development Corporation (SDC). Launched in September of 2015, Kenya RAPID is working to increase the average water coverage rate from 37% to more than 50% within five years through three strategic objectives: (1) responsive and accountable governance frameworks for providing water and pasture; (2) replicable and scalable business models for water, sanitation, hygiene, and livestock service provision; and (3) increased access to water, sanitation, and hygiene services and improved rangeland management (USAID, 2018). With over a dozen public and private partners, including the national and county ministries of water, four non-governmental organizations (NGOs), and six private companies, the Kenya RAPID program aims to strengthen institutions and facilitate coordination across sectors to improve service delivery and empower communities.
Kenya RAPID seeks to promote the consistent use of improved water and sanitation services through three activities. First, the program is supporting local financing and cost recovery through the creation of private and public partnership business models and improved asset management practices. Second, the program seeks to aid in the development of functioning management and maintenance systems through capacity-building of community-level structures and institutions. Finally, the program is promoting the protection and management of natural resources through integrated water resource management (IWRM); recharge, retention, and reuse (3R) initiatives; and the implementation of rangeland management plans (MWA, 2015). Together, these activities are aimed at building the resilience of communities located in the arid and semi-arid regions of Kenya, so that they will be able to “mitigate, recover, and adapt to major crises, such as droughts, and minor crises like rainy seasons that cut off transportation to markets” (MWA, 2015, pg. 26).

4.1.2 National Drought Management Authority

The National Drought Management Authority (NDMA) was established as a public body by the National Drought Management Authority Act of 2016. As an agency of the Government of Kenya, NDMA is “mandated to establish mechanisms which ensure that drought does not result in emergencies and that the impacts of climate change are sufficiently mitigated” (NDMA, 2015, pg. 1) As part of their mandate, NDMA is responsible for long-term planning and activities related to drought management, including disaster risk reduction, information and knowledge management, coordination of government agencies and relevant stakeholders, and rapid response to drought emergencies through the planning and implementation of contingency action plans (NDMA, 2015). The formation of the NDMA represents a significant shift in policy, “from one that relies on reacting to the effects of droughts as they arise, to one
that actively seeks to reduce vulnerability and risk through sustainable development” (NDMA, 2015, pg. 1). Intrinsic to this policy shift are the assumptions that drought disasters are avoidable—indeed, Kenya has committed to ending drought emergencies (EDE) by 2022—and that droughts are complex challenges that can only be mitigated and managed through the empowerment and increased resiliency of local institutions and communities (NDMA, 2015).

NDMA’s activities are focused in the eight arid and fifteen semi-arid counties of Kenya that are most vulnerable to drought. This vulnerability is due to multiple intersecting factors, including limited and variable precipitation; historical underinvestment in health and education services and public infrastructure; and increased land pressure due to population growth and competition for resources (NDMA, 2015). As a specialized institution tasked with coordinating across government ministries and public and private entities, NDMA activities are aimed at undermining mutually reinforcing cycles that span different policy environments. For example, investments in infrastructure and economic development aim to protect access to markets and basic services during drought events (Zommers and Singh, 2014). Similarly, investments in health and education services aim at reducing the vulnerability of households to ill-health episodes or catastrophic events (NDMA, 2015; Narayan and Petesch, 2007; Homewood et al., 2009).

NDMA’s agenda is in line with the broader transition to decentralized or devolved governance initiated in Kenya in 2010. With the establishment of county governments as a new tier of government and the creation of an Equalisation Fund to address historical underinvestment in the arid and semi-arid regions of Kenya, there is a general acknowledgement that stronger local institutions, improved security, and investments in human capital will make counties more resilient to droughts (Khaunya et al., 2015). However, unlike many of the county government structures that are vulnerable to
election cycles and disrupted by changes in administrations and political agendas, the NDMA is established as a permanent authority that is allocated funding through a National Drought Emergency Fund that receives significant financial support from the European Union (Pozzi and Oduor, 2018; EEAS, 2014, 2017). This ability to conduct planning on decadal timelines with funding that is not vulnerable to yearly budget allocations and shortfalls enables the NDMA to focus on resiliency measures, early warning systems, and longer-term investments in climate-proofed infrastructure and human capital (Zommers and Singh, 2014). Thus while NDMA’s activities are informed by a broader mandate to address historical inequalities in the development of Kenya’s arid and semi-arid regions, the nature of NDMA’s activities is highly dependent on the priorities and drought risks present in each county (Zommers and Singh, 2014). For example, NDMA officers embedded in each county work with the local county water officers to select the strategic boreholes that will receive NDMA support for parts and servicing.

4.1.3 Coupled Human and Natural Systems

In this regard, the Kenya RAPID program and NDMA provide two compelling examples of meso-level initiatives that are uniquely suited—with funding, multi-agency mandate, and a degree of political autonomy—to support counties in their efforts to avoid drought emergencies while ensuring sustainable access to water resources (Koehler et al., 2018). As depicted in Figure 4.1, the work of both of these initiatives takes place in what is often referred to as a coupled human and natural system (CHANS), a theoretical framework that describes the dynamic interplay between human alterations of the environment and the impacts on ecological systems and human well-being (Liu et al., 2007; Stevenson, 2011).

Increased drought frequency and severity have significant effects on the amount of
pasture available for grazing, the amount of available surface water, and the overall demand for groundwater resources. These changes in the natural environment are monitored through networks of individuals embedded in each of the counties, satellite imagery, and remote sensors that provide near-time assessments of water system functionality and use. The risk of drought emergency is assessed and communicated with varying efficacy throughout the different layers of governance (Haines et al., 2017). Information about need and assessments of risk are often sharpest at the local level where resources are typically scarce (Kelly et al., 2018). In contrast, information about need and assessments of risk are often wanting at the macro level where resources need to be mobilized. Thus, meso-level institutions and initiatives can play an important role in bridging these two contexts through the collection and communication of information from the local level and the coordination of response and resources from the macro-level.
To the extent that meso-level institutions are buffered from the short-term influence of election cycles, these initiatives can support long-term capacity building and policies that will reduce vulnerability to drought emergencies, like sustainable waterpoint management; surface water retention, recharge, and reuse; sustainable rangeland management; and coordination of migration and grazing (Koehler, 2018). As outlined in a recent County Diagnostic Report for Kenya published by the University of Oxford, “Risks converge in social and natural systems with the intersection of climate hazards, financial flows, operational performance, and institutional accountability,” and "information flows strengthen institutional coordination and performance” (Olago et al., 2015, pg. 18). Thus, a key finding of this report is that the use of sensors for collecting transparent, accurate, and accessible monitoring data can play a crucial role in providing accountability, unlocking new financial flows, and supporting the development of water secure institutions at scale.

An important question in this investigation is to what extent the Kenya RAPID program and NDMA activities are fostering drought resilience by improving access to sustainable groundwater services. However, since Kenya RAPID activities and coordination with NDMA are still underway, this represents only the initial assessment of how these initiatives are supporting access to groundwater services through the monitoring and maintenance of strategic boreholes in each county.

4.2 Materials and Methods

4.2.1 Study Context

Approximately 2.5 million individuals reside in the five northern counties of Kenya, and more than half of these households obtain their water from unimproved sources (KNBS / SID, 2013). Of the households that use improved sources, the vast majority
use boreholes or protected wells fitted with handpumps, while about 15% of the population use public taps from piped water systems. Almost all of the motorized pump sites from this study are located near settlements that are inhabited year-round, but a small number of the sites are used only seasonally during drought or migration periods.

As described in Haines et al. (2017), responsibility for the operation and maintenance of each water system is shared among a variety of organizations, with county water offices, sub-county water offices, water management committees, and utilities taking primary responsibility for the water systems under their purview. These entities are often supported with funding, parts, or labor by non-profits, faith-based entities like the Catholic dioceses, and drought-contingent support agencies like the NDMA.

### 4.2.2 Instrumentation and Data Management

As part of the Kenya RAPID program, program partners designed and installed remote monitoring sensors on approximately 120 motorized boreholes (SweetSense Inc, Denver, Colorado, United States). Data related to water system functionality, the approximate number of pumping hours and volume extracted per day, and the last report date for the sensor were made available online through Sweet Sense's online dashboard. In addition, IBM Research developed the Water Management as a Service Platform (WMaaSP) that will be used to collect water point data and display site functionality, use estimates, and maintenance records through an online web dashboard.

Between September 2016 and September 2018, sensors were installed on 120 motorized boreholes across the five northern counties of Kenya (see Table 4.1. Sites were selected in collaboration with county water officers, sub-county water officers, and
NDMA officials. Beginning the summer of 2016, county water officers worked with NDMA officials to identify strategic boreholes that would be prioritized for sensor installation and monitoring. Counties selected sites based on the borehole yield, the estimated human and livestock populations served, the presence of alternative water sources, and patterns of human and livestock migration during drought periods. Each list of strategic boreholes was then reviewed for security concerns before being approved by the Ministry of Water and NDMA.
Table 4.1: Sample characteristics for all data and for the drought comparison sub-study.

<table>
<thead>
<tr>
<th></th>
<th>All Data</th>
<th></th>
<th>Drought Comparison</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>strategic</td>
<td>non-strategic</td>
<td>strategic</td>
<td>non-strategic</td>
</tr>
<tr>
<td>Number of water systems</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garissa</td>
<td>15</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Isiolo</td>
<td>4</td>
<td>12</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Marsabit</td>
<td>20</td>
<td>11</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Turkana</td>
<td>16</td>
<td>19</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Wajir</td>
<td>19</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Median number of observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garissa (IQR)</td>
<td>240 (49)</td>
<td>620 (150)</td>
<td>620 (1)</td>
<td>620 (1)</td>
</tr>
<tr>
<td>Isiolo (IQR)</td>
<td>480 (3)</td>
<td>310 (220)</td>
<td>0 (0)</td>
<td>630 (130)</td>
</tr>
<tr>
<td>Marsabit (IQR)</td>
<td>350 (230)</td>
<td>360 (270)</td>
<td>640 (2)</td>
<td>640 (13)</td>
</tr>
<tr>
<td>Turkana (IQR)</td>
<td>200 (110)</td>
<td>430 (270)</td>
<td>0 (0)</td>
<td>640 (160)</td>
</tr>
<tr>
<td>Wajir (IQR)</td>
<td>360 (150)</td>
<td>600 (0)</td>
<td>610 (41)</td>
<td>600 (0)</td>
</tr>
<tr>
<td>Median households served (IQR)</td>
<td>480 (680)</td>
<td>1700 (5700)</td>
<td>420 (890)</td>
<td>2500 (6800)</td>
</tr>
<tr>
<td>Median livestock served (IQR)</td>
<td>2300 (6600)</td>
<td>0 (1100)</td>
<td>3400 (13,000)</td>
<td>0 (1200)</td>
</tr>
<tr>
<td>Power source</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generator</td>
<td>27</td>
<td>4</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Solar</td>
<td>11</td>
<td>9</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Utility</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Hybrid</td>
<td>33</td>
<td>13</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Used for filling water trucks</td>
<td>13</td>
<td>11</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Urban location</td>
<td>3</td>
<td>11</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Seasonally used pump site</td>
<td>43</td>
<td>14</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Alternative source: other pipe stand</td>
<td>29</td>
<td>23</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Alternative source: surface water</td>
<td>36</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
The sensors used in this study represent an improved version of technology used in previous studies (Thomas et al., 2013; Nagel et al., 2015; Wilson et al., 2017; Thomas et al., 2019). Sensors provided near-time feedback on borehole functionality and use. An electric current clamp recorded whether the submersible pump was running with forty-minute sampling intervals. These current signals were then used to estimate the approximate amount of time each site was being used each day. Data was logged locally on each sensor and then transmitted via cellular and satellite networks to remote servers for analysis. Each day an R script estimated the approximate number of hours of use for each site. Where flow rate data were available for specific pump sites, the length of use was used to estimate the amount of water extracted at each site. Sensor up-time (percentage of days of reported sensor functionality) across the 120 sites averaged 96.9% (SD = 6.57%). Site up-time (percentage of days of at least some recorded use) averaged 59.5% across sites (SD = 27.5%). The use of artificial intelligence and expert-informed thresholds for determining site status (i.e., sensor malfunction, water system malfunction, seasonal disuse, or fully functional system) will be described in a forthcoming publication. At twenty-four of the sites, the electrical clamp was cut or removed causing the sensor to record that the pump was not being used. Representing approximately 5,000 observations or about 5% of the total dataset, these data were removed prior to conducting the analysis.

A summary of pump site statuses, last report date, and the approximate volume extracted each week were made available through the Sweet Sense dashboard. Similar information was also available through the WMaaSP dashboard. Sensor installation and removal records as well as repair reports were logged using the mWater mobile and online applications (Feighery et al., 2015). In addition, asset inventory surveys were completed by Sweet Sense field officers and county water officers for each site to record information about water use, water quality, borehole characteristics, power
source and characteristics, and water point management.

Finally, precipitation was estimated using monthly Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) v2.0 according to the methodology described in Thomas et al. (2019). Gridded raster precipitation files with a resolution of 5 km were used to estimate rainfall in the proximity of each water system. However, these precipitation estimates were not used in this analysis as we found that seasonality (rainy vs. dry season) was a better predictor of use than daily precipitation values.

4.2.3 Statistical Analysis

To characterize the factors that influenced motorized borehole use, we examined sensor and water point data between September, 2016 and January, 2019. Sensors were installed and replaced over the course of this study, resulting in an unbalanced number of observations for each pump site. Given the nested structure of the data, we used a two-level multilevel linear regression model with random intercepts and an AR-1 correlation structure to account for autocorrelation and non-independence of the errors. Individual observations were nested within each site, and one time-varying predictor was incorporated: a rainy vs. dry season indicator, where the rainy season corresponded to the primary and secondary rainfall periods of March through May and September through November (Camberlin, 2018). In addition, several site-level predictors were examined, including borehole status (strategic vs. non-strategic), number of households served, number of livestock served, power source (hybrid, solar, generator, or utility power), alternative sources of water (a surface water source or a different water tap), whether the water point was used for filling water trucks, and urban vs. rural location.

As there was almost half the amount of precipitation in Northern Kenya during
2017 compared to 2018, we also conducted a subgroup analysis to compare overall site utilization under two different drought conditions. In 2017, a state of emergency was declared due to reduced rainfall during the primary rainy season (March through May), catalyzing significant resources and support from organizations like the NDMA, non-profits, county governments, and the federal government. Since only twenty-five sensors were installed at the beginning of the 2017 drought season—eight of those sensors were installed on strategic boreholes—we compared how utilization differed between 2017 and 2018 for this subset of sites. All statistical analyses were conducted using R, and the multilevel models were evaluated using the nlme package (Pinheiro et al., 2018).

4.3 Results

Average precipitation across the pump sites was 171 mm in 2017, compared to 326 mm in 2018. As seen in Figure 4.2, there was considerable variation within each county between the two years. Isiolo received the least precipitation in 2017 at 111mm, while Marsabit received the most precipitation at 298mm. In 2018, Isiolo received the most precipitation at 523mm, while Turkana received the least precipitation at 203mm. Water systems were used on average 7.1 hours in 2017 versus 5.4 hours in 2018. Average utilization also varied considerably by county, with all counties except Turkana averaging greater use in 2017 than in 2018. Garissa had the lowest average use both years (3.6 and 3.3 hours, respectively), while Wajir had the greatest use in 2017 (9.5 hours) and Turkana had the greatest use in 2018 (8.0 hours). Based on estimated flow rates from individual sites, a daily average of 142 m$^3$ of water was extracted from each site in 2017, compared to 98 m$^3$ in 2018.

In the null model with no covariates and intercepts allowed to vary across individual pump sites, there was considerable variation both within and between sites. The
intraclass correlation coefficient, which provides a measure of the amount of variance attributable to differences between sites, was 45.3%. Variability across time within sites ($\sigma^2 = 4.26$) was slightly greater than variability between sites ($\tau^2 = 5.15$).

As seen in Figure 4.3, water system use was strongly tied to precipitation, although the influence of precipitation on water system use varied considerably by county. In counties with greater variability in use like Marsabit and Garissa, there was significant ramping behavior whereby groundwater demand would increase significantly during the dry season and then drop precipitously after rainfall events. In counties with more consistent water system use like Turkana and Isiolo, there were more systems that were run for predictable intervals (e.g., 12- or 24-hour intervals). As a result, water system use in these counties was more influenced by system functionality than precipitation events.
Figure 4.3: Left axis: average daily water system use for each county from 2017 through 2018. The red line indicates average daily use across all boreholes for each county, whereas the blue and pink dots show the individual observations for strategic and non-strategic boreholes, respectively. Right axis: the blue line describes the average daily precipitation across all pump sites for each county. The dark grey region demarcates the primary rainy season (March through May) and the light grey region demarcates the secondary rainy season (September through November).

As seen in Table 4.2, there were a variety of site-level characteristics that were strong predictors of water system utilization. As a fixed effect, the county each
The site was located in was a significant predictor of utilization, with all counties except Isiolo recording greater average daily use than the reference county of Garissa. The type of power system was also a significant predictor of use, with hybrid and utility systems being used 1.1 and 3.5 hours more than sites with generators ($p = 0.0043$ and $p < 0.001$). Surprisingly, water systems powered by solar were used almost an hour less per day compared to sites with generators; however the difference was not statistically significant ($p = 0.0597$).

<table>
<thead>
<tr>
<th>Table 4.2: Multilevel model results for all data.</th>
<th>Coefficient (SE)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.95 (0.795)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>County: Garissa (reference)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>County: Isiolo</td>
<td>-1.94 (0.686)</td>
<td>0.0058</td>
</tr>
<tr>
<td>County: Marsabit</td>
<td>0.514 (0.592)</td>
<td>0.388</td>
</tr>
<tr>
<td>County: Turkana</td>
<td>2.33 (0.690)</td>
<td>0.0011</td>
</tr>
<tr>
<td>County: Wajir</td>
<td>2.39 (0.558)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Strategic Borehole Status</td>
<td>-1.33 (0.384)</td>
<td>0.0009</td>
</tr>
<tr>
<td>Households Served (1000s)</td>
<td>-0.195 (0.0541)</td>
<td>0.0005</td>
</tr>
<tr>
<td>Livestock Served (1000s)</td>
<td>-0.0343 (0.0149)</td>
<td>0.0242</td>
</tr>
<tr>
<td>Power: Generator (reference)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Power: Hybrid</td>
<td>1.11 (0.375)</td>
<td>0.0043</td>
</tr>
<tr>
<td>Power: Solar</td>
<td>-0.959 (0.516)</td>
<td>0.0667</td>
</tr>
<tr>
<td>Power: Utility</td>
<td>3.56 (0.666)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Seasonal Pump</td>
<td>0.685 (0.319)</td>
<td>0.0349</td>
</tr>
<tr>
<td>Alt. Surface Water Source</td>
<td>-0.982 (0.337)</td>
<td>0.0091</td>
</tr>
<tr>
<td>Rainy Season</td>
<td>-0.389 (0.126)</td>
<td>0.0021</td>
</tr>
<tr>
<td>Water Trucking</td>
<td>3.03 (0.348)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Urban Location</td>
<td>3.23 (0.578)</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

The number of households and livestock being served by each water system was also statistically significant ($p = 0.005$ and $p = 0.0242$), although the effect size per thousand was small (-0.195 and -0.0343). However, sites used for filling water trucks were used on average three hours more compared to sites not used for water trucking ($p < 0.001$), and urban sites were used 3.25 hours more than rural sites ($p < 0.001$). Sites where the operator reported using the water system less during the
rainy season averaged slightly more utilization on average compared to non-seasonal sites ($0.713, p = 0.0278$). However, in general sites were used 0.383 hours less during the rainy season compared to the dry season ($p = 0.0024$). Similarly, sites that reported using a surface water source when the water system was non-functional were used almost an hour less per day on average compared to sites that reported using a different water tap ($p = 0.0128$).

A surprising finding is that sites designated as strategic boreholes were used almost two hours less on average compared to non-strategic boreholes ($p = 0.0010$), even when controlling for all the fixed effects listed above. In order to better understand why strategic boreholes were used less, we constructed two additional models to explore the interactions between borehole status, county activities, and drought conditions.

Figure 4.4: A: Average marginal effect of strategic borehole status on hours of utilization for each county. B: Average marginal effect of strategic borehole status on hours of utilization for the 2017 and 2018 drought seasons.
In the first exploratory model, we found that interactions between borehole status and county were only significant for Isiolo and Wajir. This is further demonstrated in Figure 4.4, where the average reduction in water system use for strategic boreholes is greater for these two counties. However, it is important to note that borehole statuses are severely unbalanced in Wajir and Isiolo because there is only one non-strategic borehole in Wajir and only four strategic boreholes in Isiolo.

In the second exploratory model we looked at differences in water system use when controlling for drought conditions. As noted previously, there was significantly less precipitation in 2017 compared to 2018. One consequence is that the NDMA was more active in each of the counties in 2017, not only assessing drought risk and coordinating execution of each county’s contingency action plan, but also providing significant assistance through water trucking, fuel subsidies, water equipment, and support to emergency repair teams.

When the 2017 drought began in June, there were only twenty-five sensors installed on water systems, and only eight of those water systems were designated as strategic boreholes. Therefore, we looked at changes in use for those twenty-five water systems while controlling for the 2017 and 2018 dry season (June through September). However, all the predictors that were significant in the full model were no longer significant, but water systems were used about 2.69 hours more in 2017 compared to 2018 ($SE = 0.645, p < 0.001$). Further, when borehole status and drought year were interacted in a separate model the interaction was significant ($\beta = -4.10, SE = 1.30, p = 0.0016$). As seen in Figure 4.4, there is a significant difference in the average marginal effect of borehole status when interacted with drought year. Whereas the seventeen non-strategic boreholes were used approximately the same amount of time during both dry seasons, strategic borehole use dropped from 6.5 hours per day in 2017 to less than three hours per day in 2018.
4.4 Discussion

Fostering resilience to drought emergencies is a complex task, and the effort to improve sustainable water services describes just one facet of a broader movement to build capacity and reduce vulnerability in Kenya’s northern counties. However, sustainable water services have a disproportionate impact on the severity of drought disasters, as the inability to access safe or improved water services has ramifications that are felt rapidly and touch all other aspects of community life in rural regions (MWA, 2015). While traditional drought strategies have focused on food supplies and getting humanitarian aid to the places where it is needed in a timely manner, recent strategies have shifted their focus to preventing drought emergencies by fostering resilience, building capacity, and developing early warning and rapid response systems to avoid the worst effects of drought emergencies: inter-communal conflict over scarce resources, the destruction of community assets, and the loss of human lives (Fitzgibbon and Crosskey, 2013).

This investigation builds on previous studies that have shown how specialized institutions equipped with continuous monitoring technology can significantly improve the sustainability of water services in developing communities (Thomson et al., 2014; Nagel et al., 2015; Koehler et al., 2018). For example, Nagel et al. (2015) conducted an experiment where functionality increased significantly (91%) while time to repair decreased (21 days) using sensor-informed maintenance compared to circuit-rider maintenance (73% functionality and 57 days to repair) or nominal maintenance methods (68% functionality and 152 days to repair). Similarly, Thomson et al. (2014) used handpump sensors and a sensor-informed maintenance model to raise pump functionality to 98%, with 89% of repairs being completed within five days.

However, both studies are quick to point out that the data from sensors alone are
not sufficient to improve the operation and maintenance of water systems. Researchers at the University of Oxford have contributed significantly to the effort to articulate how smart monitoring must be combined with professional maintenance services, sustainable finance, and coordination in order to provide sustainable water services (Thomson and Koehler, 2016). Koehler et al. extend this analysis further in their articulation of the need for pluralist institutions that “facilitate risk management through improved information flows, sustainable finance and reliable maintenance” (2018, pg. 272). These pluralist institutions ideally facilitate collaboration between the market, bureaucratic, and community perspectives that must be represented in the effort to reconcile conflicting values and manage a common pool resource like water (Koehler et al., 2018). While the FundiFix model and Kenya RAPID programs have been implemented as external interventions, the NDMA provides a compelling example of a pluralist, meso-level institution that is able to “reduce operational and financial inefficiencies by pooling finances and operating at scale” while channeling resources from the federal budget to the communities and contexts where and when they are needed (Koehler, 2018, pg. 275).

The NDMA’s ability to support each county’s waterpoint management is dependent on the drought conditions that catalyze resources. Whereas the drought in 2017 provided additional resources for fuel, maintenance, and water equipment, resulting in greater overall use of strategic boreholes, the lack of drought in 2018 resulted in diminished overall use of boreholes, particularly strategic boreholes. Based on interviews with field officers and an examination of repair records, one reason cited for decreased use of strategic boreholes in 2018 compared to 2017 is that strategic boreholes in 2018 were disproportionately affected by malfunctions and prolonged repairs. It is not clear if the number of malfunctions experienced between 2017 and 2018 is different, or if there were simply fewer resources in 2018 to conduct repairs in
a timely fashion. In addition to the lack of resources normally triggered by drought conditions, the contested presidential election and expensive re-vote in 2017 could have contributed to additional budget shortfall for county water offices (The Carter Center, 2018). All five counties involved in this study reported limited resources for paying personnel or mobilizing vehicles or repairs during the months of June through August when Kenya was entering a new annual financial cycle.

It is also possible that water systems at strategic boreholes were used less in 2018 simply because there was less demand. This investigation also builds on the growing literature demonstrating how ground water demand varies significantly depending on precipitation events and the availability of surface water sources (Thomson et al., 2018; Kelly et al., 2018; Thomas et al., 2019). As outlined in Thomson et al. (2018), the implications of fluctuating demand for basic water services are significant because previous studies have suggested that high adherence to safe water is necessary to realize sustained health benefits (Brown and Clasen, 2012; Enger et al., 2012; Hunter et al., 2009; Howard et al., 2006). As discussed in Thomson et al., the apparent tendency of households to switch from improved to unimproved water sources based on surface water availability merits further investigation. Much like stove stacking with cookstoves (Masera et al., 2000), water source stacking challenges the assumption that keeping waterpoints functional will be sufficient to decrease the burden of disease from water-borne illnesses (Thomson et al., 2018). The apparent elasticity of demand also suggests that operation and maintenance be focused where and when demand is least elastic: during dry periods, in areas with limited surface water availability, or in areas where demand for clean water is already high.

There are several limitations of this study. For example, the fact that we measured water system use instead of functionality means that we were unable to reliably distinguish when water systems were broken from when water systems were simply
not being used. Also, we observed that the definition of strategic borehole was applied differently in each county. While this preserves each county’s prerogative for prioritizing particular boreholes, it also makes comparisons across counties difficult. For example, in Isiolo there was initially difficulty identifying which boreholes were designated as strategic due to discrepancies between the NDMA officer and the county water office. Of the four that were labelled strategic, many were closed down during the year and only mobilized during drought conditions. In contrast, most of the strategic boreholes in Turkana were used all year round, with variable usage patterns depending on the extent of human and livestock migration during the dry season. Also, at the time of writing a process was still being formalized for operationalizing these data to inform maintenance activities in each county. Also, it is worth noting that the number of water points with sensors represented a small fraction of the total number of motorized and non-motorized boreholes in each county. Due to the evolving nature of the online dashboards and the inchoate processes for integrating sensor data into existing operation and maintenance activities, the effect of having access to information about each pump site on overall site use and functionality was not considered in this phase of the study.

Finally, much of this investigation has focused on the unique role that meso-level institutions and initiatives can play in reducing the vulnerability of communities in arid regions to drought disasters. While we have emphasized the value of using continuous waterpoint monitoring data to inform and mobilize support for the operation and maintenance activities of local county water offices, we acknowledge that water security cannot be reduced to waterpoint management. Given that these activities take place on a complex and adaptive ecological and social landscape, it will be important to explore how funding and resources for waterpoint management can be balanced with the other initiatives being undertaken by each county to increase drought re-
siliency. Since all of these initiatives are tightly interlinked—for example, surface water catchments can increase surface water availability and groundwater recharge, and sustainable rangeland management can reduce the need for human and livestock migration—it will be important to consider whether the sum of these activities is greater than any of their individual contributions.

Acknowledgements

The data utilized in this publication were made possible through support provided by the U.S. Agency for International Development, under the terms of Cooperative Agreement No. AID-615-A-15-00008. Additional support was provided by the Link Foundation and the National Science Foundation IGERT Grant #0966376: “Sustaining Ecosystem Services to Support Rapidly Urbanizing Areas.” Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Agency for International Development or the National Science Foundation. Our partners include the Kenya National Drought Management Authority, the Millennium Water Alliance, Catholic Relief Services, World Vision, mWater, IRC, CARE, IRC Wash, IBM Research, SweetSense Inc., the US-AID Program, and the county governments of Turkana, Isiolo, Wajir, Garissa and Marsabit in Kenya.
References


J. Feighery, R. Smith, C. Grassick, and A. Feighery. mWater: a free and open-access platform for water data sharing and collaboration. *Open Water Journal*, 3(1), 2015. URL [https://scholarsarchive.byu.edu/openwater/vol3/iss1/7](https://scholarsarchive.byu.edu/openwater/vol3/iss1/7).


Chapter 5

Conclusion

5.1 Weaving It Together

The field of information and communications technologies for development (ICT4D) is broad, and this investigation represents one small contribution to the larger effort to improve the ways that progress is measured with respect to global water, sanitation, and hygiene development goals. While the preceding three chapters were framed around specific research efforts being undertaken by the Sustainable Water, Energy, and Environmental Technologies Lab at Portland State University, one of the critical features that links all three efforts is the consideration of feedback and how it can be effectively integrated into development projects to improve outcomes and impacts. Figure 5.1 provides a conceptual diagram of the ways that single- and double-loop feedback from different data sources can be incorporated into development projects to monitor key performance indicators and provide more effective measurement of specific outcomes and impacts. The layers of feedback from each investigation are highlighted in red with more detailed diagrams provided below.

Given the complexity of the diagram, descriptions of the feedback configurations that were explored in the previous three chapters are included here:

1. Chapter 2: Overview of Sanitation and Hygiene Measurement Methodologies and Technologies
Figure 5.1: Conceptual diagram of the sources, measures, and outputs of monitoring and evaluation methodologies. Each research effort is framed in red.

(a) Presented causal inference models in relation to indicators used to measure access to sanitation and hygiene services

(b) Reviewed literature pertaining to innovative methodologies and technologies being used to measure access to and use of sanitation and hygiene services


2. Chapter 3: Dynamic Management of Sanitation Services in Nairobi, Kenya

(a) Partnered with a social enterprise that provides safe sanitation solutions to individuals living in informal settlements

(b) Used motion sensor data, manually recorded weight data, and machine learning algorithms to create a dynamic schedule for latrine servicing

(c) Project was adaptively managed by integrating feedback from field officers and waste collectors to tune the sensitivity of the algorithm and more effectively identify which latrines could be skipped each day


(g) Oral presentation and poster: Turman-Bryant, Nick. (2018) Predicting overflow: A novel application of latrine sensors and ensemble learning for
Chapter 4: Improving Drought Resilience for Strategic Boreholes in Arid Counties in Kenya

(a) Partnered with a large international development agency, several public and private organizations, a drought management authority, and different ministries within the Kenyan federal and county governments to improve drought resilience for communities residing in five arid counties

(b) Used electrical current sensors and maintenance records to estimate the functionality and utilization of different water systems and to monitor operation and maintenance activities

(c) Explored the unique roles that remote sensors and meso-level institutions can play in mediating information and resources between macro-level organizations (e.g., federal government or international development agencies) and micro-level communities


In practice, the application of this framework was different for each research effort. For example, Figure 5.2 maps the above feedback diagram to the second research effort exploring improved sanitation service delivery. Given that there was only one managing body (Sanergy) and the monitoring and evaluation was limited to project outputs and outcomes, the investigation was primarily concerned with how near-time
feedback could improve sanitation service delivery. However, the near-time feedback from sensors and implementers also provided an opportunity to explore how the project design and management could be adapted during implementation to improve project outcomes.

Similarly, Figure 5.3 provides a visual schematic of how sensor data were used to monitor the operation and maintenance of boreholes in Northern Kenya as part of a larger effort to improve drought resilience in arid regions. Although sensor data and estimates of water system functionality and use were made available to county water officers through an online dashboard, the five counties were still formalizing processes for tracking water systems to inform maintenance activities. As a result, this investigation represents more of a baseline analysis of how variability in water system
use is explained by factors related to the water system (e.g., number of households or livestock served, power source, location, etc.) as well as exogenous factors (e.g., time of year, amount of precipitation, availability of surface water sources, etc.). In particular, this investigation used feedback from sensor data and maintenance records to explore how utilization differed between boreholes that counties had designated as strategic—Ending Drought Emergency (EDE) boreholes—compared to non-strategic boreholes.

Thus, while the second chapter provided an evaluation of the appropriateness and accuracy of different indicators that are used to measure access to services, the third and fourth chapters focused on how the information and insights extracted from complex settings can be applied to improve desired outcomes. For example, the second chapter asked, “What are the different facets of access, and how can we accurately and appropriately measure them? How are new technologies and methodologies changing
the ways that we measure progress on development goals?” In contrast, the third and fourth chapters were more concerned with how knowledge is derived from data and then translated into action. In a complex adaptive system like an informal settlement in Nairobi, the third chapter explored how data from simple technologies can be combined with sophisticated machine learning methodologies to improve the reliability and cost-effectiveness of sanitation services. Similarly, the fourth chapter explored how data from remote sensors can be used to derive insights and inform decisions in a coupled human and natural systems setting to improve resilience to drought disasters. In addition, this investigation explored how institutional characteristics play a role in facilitating the management of and response to a complex and adaptive phenomenon like a drought.

5.2 What About the Constraints of ICT4D?

Much of the preceding chapters was focused on the ways that information and communication technologies can be used to improve the delivery of water, sanitation, and hygiene services. While each of these investigations highlighted specific questions that were related to the work being done by Sweet Sense, Inc. and the research being conducted by the SweetLab, each chapter also addressed particular shortcomings or opportunities for further investigation. However, many of the constraints described in these investigations were particular to the methodologies or data sources used in the studies. In an effort to provide more perspective on the potential constraints or pitfalls associated with using information and communication technologies in development, the following section explores critiques of ICT4D from the literature that can shed light on the preceding chapters and point to opportunities for further research.
5.2.1 ICTs and the Poor as Consumer

Traditional development discourse has focused on the poor as passive recipients of aid, with an emphasis on “providing aid, investment, support, institutions, and capacity to improve social and economic conditions” (Kuriyan et al., 2012, pg. 2). In contrast, the more recent shift toward focusing on the poor-as-consumer conceptualizes the poor as “active agents within a consumer market, with legitimate spending needs, demands, desires, and constraints to be addressed” (Kuriyan et al., 2012, pg. 2). For example, Prahalad and Hart suggests that “[t]he real source of market promise is not the wealthy few in the developing world, or even the emerging middle-income consumers: It is the billions of aspiring poor who are joining the market economy for the first time” (2002, pg. 1). In their conception, the poor are not able to participate in the global market economy due to the extreme inequity of wealth distribution and constraints that have locked them out of the formal economy (e.g., lack of credit history, lack of formal titles or deeds to property, lack of education, etc.). Similar to Jeffery Sachs’ articulation of the poverty trap that keeps the poor from reaching the first rung of the development ladder, this conception of the poor-as-consumer identifies structural constraints that contribute the social reproduction of poverty and inequality across generations (Sachs, 2005). Information and communication technologies are often viewed as one of the most important mechanisms for connecting with poor with the marketplace, not only because of the widespread demand for mobile and technology devices, but also because these devices are assumed to improve connectivity and access to information that can result in increased productivity, educational benefits, and social networking (Kuriyan et al., 2008; Nederveen Pieterse, 2010).

However, more recently there has been push-back against using the poor-as-consumer framing as the primary driver of poverty alleviation efforts. Indeed, while the poor-as-consumer framing provides a needed correction to the former conception
of the poor as passive recipients of donor-based aid, the Bottom of the Pyramid (BOP) paradigm can swing to the opposite extreme. For example, Prahalad suggests that multinational corporations (MNCs) are uniquely positioned to serve BOP customers due to the inefficiencies of governments and the limited resources and expertise of small- and medium-sized businesses (Prahalad and Hart, 2002). However, a short review of MNC involvement in emerging economies would reveal that the impact on the poor and the environment has been mixed at best (Watts, 2005; Pearson, 1985; Zelman, 1990). Firm believers in the power of MNCs and markets to lift the BOP out of poverty while making a profit perhaps underestimate the role that national governments and multilateral institutions must play in regulating market mechanisms to account for market failures (Easterly, 2015).

Beyond the problems of implementation, there are also specific shortcomings associated with the poor-as-consumer perspective as it is popularly conceived. For example, Kuriyan et al. suggest that treating the poor as homogeneous “glosses over critical differences of position, power, behavior, and incentives within these categories, or the relations between low-income people and the state or private sector” (Kuriyan et al., 2012, pg. 2). Furthermore, Kuriyan et al. argue that casting the poor as consumers is another means of “deemphasizing their political and cultural embeddedness” (Kuriyan et al., 2012, pg. 3). Finally, the conception of poor-as-consumer can underemphasize the social and psychological factors driving behavior in an attempt to rationalize consumer behavior based on price sensitivity or efficient resource allocation. Instead, they suggest that low-income groups should not necessarily be targeted specifically, which could result in their feeling isolated, but rather focus marketing campaigns on a “broader swath of society” through which the poor can construct aspirational identities (Kuriyan et al., 2012, pg. 8). Roy et al. also suggest that casting the poor as consumers reframes consumption “not as the prob-
lem but as the solution... Far from connecting the consumption of luxury goods in the United States to the exploitation of the labor, land, and natural resources of the global South, [Western] consumers...are given a way to satisfy both their material desires and their consciences at the same time” (2016, pg. 151).

As discussed in Kuriyan et al. (2008), there is an implicit tension when attempting to improve the well-being and livelihood of the poor through public-private partnerships. On the one hand, governments and non-governmental organizations (NGOs) want to achieve specific social development goals (e.g., increased literacy, education, improved health, increased productivity or income, etc.) while private-sector interests are interested in commercial profitability. In their article, “Information and Communication Technologies for Development: The Bottom of the Pyramid Model in Practice,” Kuriyan et al. document the Indian government’s attempt to meet these twin goals of social development and financial sustainability through the promotion of education courses offered through computer service kiosks in Kerala (2008). While Kuriyan et al. tease out the tensions inherent in such an effort, they also ask a broader and more provocative question—to what extent are social development needs even being addressed through the dissemination of information and communication technologies?—calling out the almost implicit assumption that the adoption and use of ICTs by consumers at the bottom of the pyramid is automatically beneficial.

5.2.2 ICTs and the ‘Rendering Technical’ of Development Work

The casting of the poor as consumers is tied to a broader critique of ICT4D as supporting a neoliberal development paradigm that marginalizes the role of the state. In her book, *Metrics: What Counts in Global Health*, Adams questions whether the imposition of global health agendas and the orchestration of interventions and investments by multilateral institutions and NGOs can compromise or conflict with the
national interests of individual nation states (2016). Similarly, Roy argues that there is a lack of recognition that the “problems” that have been rendered apolitical or neutral stem from the underlying persistence of injustice and inequality (2016, pg. 16). For example, reframing poverty as inequality has significant political implications. Whereas poverty is abstracted, apolitical, decontextualized, and renders the poor as passive victims, inequality implies injustice and an opportunity for empowerment (Roy, 2016). This problematization of political inequalities also empowers external trustees who have expert knowledge with the responsibility to enact the policies that they feel are most appropriate or beneficial.

Neoliberal development paradigms and policies are a favorite target for those who question the practice of “rendering technical” the underdevelopment or poverty of emerging economies (Li, 2007; Ferguson and Lohmann, 1994). This could be because, according to Stiglitz, neoliberal policies attempt to reduce the development of a country to the problem of liberalization, privatization, and macro-economic stabilization (Stiglitz, 2001). It is also true that many see economic development as the key determinant of social development, although this idea is much less popular after several decades of structural adjustments and strong neoliberal policies being imposed on developing economies with little improvement in social welfare (Polanyi, 2001; Thomas, 2000). For this reason, while GDP per capita is still a popular metric of development, a variety of other metrics have been developed to capture the diversity of needs within emerging economies (e.g., HDI, MDGs, SDGs, etc., Thomas, 2000).

Still, there is a persistent assumption, originally identified by Polanyi, that the economy can be disembedded from the broader functioning of society (Block, 2001, pg. xxv). Moreover, the practices and policies of development can often be at odds with the functioning or sovereignty of the state, particularly when the state is seen
as being a principle contributor to societal problems due to inefficiency, corruption, nepotism, etc. (Prahalad and Hart, 2002; Adams, 2016). For this reason, there is a strong incentive to problematize the social issues found in emerging economies (“identify deficiencies that need to be rectified”) and render them technical (i.e., introduce experts as trustees and depoliticize the problem, Li, 2007, pg. 7). Adams also provides a thorough critique of the ways in which scientific metrics (like randomized controlled trials) indirectly become the micropractices of neoliberalism by enabling the fiscalization of life through esoteric measures (e.g., QALY, DALY, or HALY) and adapting the problems of developing countries “so that they work within capitalism’s terms and limits and so that they serve the global architecture of neoliberal debt and profit economies” (2016, pg. 45).

In this way, the belief in the “natural” or self-regulating tendency of the market ignores the political dimensions of the policies being imposed (in particular, who benefits and whose interests are being represented). It also empowers external trustees who have expert knowledge with the responsibility to enact the policies that they feel are most appropriate or beneficial. It characterizes problems in purely economic terms, often forcing countries to sacrifice the well-being of their citizens and to relinquish the freedom of their democracy to enact the policies (Polanyi, 2001; Adams, 2016). Finally, blind faith in neoliberal policies ultimately ignores the way that the practices of the global economic system actually contribute to the impoverishment of the country by saddling them with high interest loans or bailing out private investors instead of providing for bankruptcy options (Stiglitz, 2001).

For some, ICT4D plays a critical role in the neoliberal expansion of digital capitalism because it allows ICT proponents to frame ICT market support and development as critical components in the wider effort to close a ‘digital divide’ between the technological ‘haves’ and ‘have-nots’. However, critical development theorists
like Jan Nederveen Pieterse question the technocratic assumption that the spread of technology and market forces are synonymous with development (2010). Nederveen Pieterse provides a list of potential pitfalls of the ICT4D paradigm, including: (1) ICT4D policies often conflate human capital outcomes like education services with digital inclusion and literacy; (2) ICTs are often designed for western markets, they privilege western content, and they presuppose western legal norms; (3) ICT development is often too technical or capital intensive to be participatory and thus often depends on ‘development from above’; and (4) ICTs can become a means of control and exclusion through surveillance and corporate resistance to open source software (Nederveen Pieterse, 2010, pg. 309-312, 315). Of course, even strong critics of ICT4D do not dismiss the immense potential for empowerment and connectivity that ICTs can facilitate. Instead, Nederveen Pieterse suggests that “ICT4D is a prism in which profiles of neoliberal globalization are refracted. It stands at the crossroads of today’s major forces in private, public and social spheres: telecoms, international institutions, states and civil society groups and cyber activists” (Nederveen Pieterse, 2010, pg. 314).

5.3 The Evolution of Evaluation

One of the contributions of this investigation is an in-depth analysis of the methodologies that can be used to distill and operationalize insights from intensive longitudinal data. For example, the second chapter presented a discussion of how causal modeling can be used to understand inferential uncertainty and inform indicator selection for measuring progress on sanitation and hygiene targets. Similarly, chapter three used a sophisticated machine learning algorithm to predict which latrines could be skipped based on near-time sensor and weight measurement data. Finally, chapter four used a multilevel model to compare utilization of strategic and non-strategic water systems
to better understand the factors that contribute to drought resilience in arid regions of Kenya.

In some respects these three chapters highlight the ways that evaluation science is being transformed by the influx of Big Data from IoT and mobile devices (Biggs et al., 2016; Khoury and Ioannidis, 2014). Part of this transformation is due to the fact that large datasets with higher resolutions allow researchers to make inferences and derive insights more rapidly than traditional clinical trials that used repeated measures taken at less frequent intervals (Walls et al., 2017). Another aspect of this transformation is the blurring of lines between the different phases of research. For example, Figure 5.4 displays the research classifications used in translational science to describe the “the process of turning observations in the laboratory, clinic, and community into interventions that improve the health of individuals and the public—from diagnostics and therapeutics to medical procedures and behavioral changes” (NCATS, 2019). These phases are assumed to progress from basic research in a clinical or laboratory setting to actual dissemination and implementation in communities at scale. Along the way, it is assumed that the research undergoes proof-of-concept studies, efficacy trials, and effectiveness trials to ensure that the results observed in controlled and closely monitored laboratory or clinical settings translate to the field where exogenous factors are more difficult to control and behaviors are more difficult to monitor.

However there is a growing appreciation of the difficulties inherent in transitioning from laboratory or clinical settings to field settings, what is often referred to as the “efficacy-effectiveness gap” (Nordon et al., 2016). This discrepancy between efficacy and effectiveness has motivated some researchers to argue for more rapid incorporation and closer monitoring of real world data to provide evidence of effectiveness for therapies and practices being implemented in the field (see Figure 5.5, Katkade et al., 2018). The need for more effectiveness evidence has also given rise to the
nascent field of implementation science, which aims to “bridge the ‘know-do gap’ by applying methodologically rigorous approaches to 1) generating and synthesizing population-level evidence within context, 2) testing interventions informed by contextual evidence, 3) translating appropriate findings into practice and, 4) continuing the evidence generation cycle” (Means et al., 2016). The third and fourth chapters in this investigation provided salient examples of this efficacy-effectiveness gap. In the third chapter, there was a significant difference between the simulated results of the predictive models and the actual results when a dynamic model was deployed, and the gap was primarily attributable to challenges experienced in implementation. Similarly, in the fourth chapter we observed that groundwater demand decreased significantly after precipitation events, challenging the assumption that individuals and households will use improved drinking water sources as long as they are functional. In this way, implementation science research is providing an important corrective to a translation paradigm that has focused more on efficacy than effectiveness by highlighting the context and behaviors that are necessary for a given intervention to be

---

**Figure 5.4: Phases of research in a translational science schema.** T1TR and T2TR stand for Type 1 and Type 2 Translational Research (ICTR, 2019).
The challenge is that global health interventions rarely follow the trajectory described by the translational science diagram (see Figure 5.4). Cookstoves, water filters, malaria nets, latrines, or nutritional supplements that perform well in laboratory setting are often implemented at scale without a clear understanding of the thresholds or adherence to treatment that are necessary to have a significant health impact. Similar to the quality of evidence graphic in Figure 5.6, interventions are often deployed based on expert opinion or anecdotal evidence that clean water, air, sanitation, and hygiene are necessary for improved household health, but it is not always evident how clean or under what circumstances these interventions are actually effective. In this regard, global health interventions are at a triple-disadvantage
because they are often lacking in efficacy, effectiveness, and implementation evidence. Health impacts have often been modeled based on dose-response curves that can estimate risk reduction based on reduced exposure to pathogens or poor air quality (Brown and Clasen, 2012; Havelaar and Melse, 2003; WHO, 2011; Johnson and Chiang, 2015), but there is still much efficacy evidence that is lacking. For example, with regard to air quality it is generally agreed that less pollution is better, but it is still unclear which thresholds of air quality—both in frequency and in magnitude of exposure—are associated with significant health benefits. Similarly, safely managed sanitation and basic hygiene services are recognized as vital interventions for reducing exposure to enteric pathogens in a household environment, but there is still little evidence about which technologies and management behaviors are most effective for reducing risks for households and communities (Turman-Bryant et al., 2018). Even when safe thresholds have been declared by the World Health Organization (e.g., no detectable organisms per 100 mL of water or less than 10 $\mu g/m^3$ annual exposure to PM$_{2.5}$, WHO, 2011, 2006), there is still a lot of uncertainty surrounding the environmental factors or types of behaviors that are necessary for effective implementation—for example, how exclusively an improved cookstove must be used or the level of adherence to safe drinking water sources (Brown and Clasen, 2012; Shankar et al., 2014). Similarly, there is growing evidence that health benefits are tied to interactions between interventions, requiring a more integrated perspective of household health (Clasen and Smith, 2019). For example, respiratory infections due to poor air quality may increase a person’s susceptibility to enteric infections due to unsafe water, sanitation, or hygiene conditions (Ashraf et al., 2013).

In the context of global health where effectiveness and implementation evidence are still being established, implementation science has codified the concept of hybrid effectiveness-implementation designs that focus on both the implementation and
Figure 5.6: “General hierarchy of study types” (Katkade et al., 2018).

health outcomes (Bauer et al., 2015). As seen in Figure 5.7, these hybrid designs aim to address the efficacy-effectiveness gap by formalizing methodologies for assessing how well an intervention is implemented while still evaluating the impact of the intervention on a desired health outcome. In this way, implementation science provides a disciplinary bridge between the worlds of monitoring and evaluation by insisting that continuous monitoring data be incorporated into implementation processes in order to improve the overall outcomes and impacts of an intervention (Bauer et al., 2015).

As discussed in the introduction, monitoring and evaluation in development typically imply a trade-off between actionable learning and generalizable impact, where ex-post evaluations can be constrained in their ability to respond to insights gained during implementation. However, the combination of hybrid experiment designs with technologies that provide near-time feedback at higher resolution are enabling new opportunities for adaptive monitoring and evaluation of development interventions like the studies described in chapters three and four. The intensive longitudinal data
generated by remote sensors can then be analyzed using multilevel models or linear splines that preserve the complexity of specific site characteristics when comparing adaptations in the intervention without reducing the data to a simple comparison of aggregated outcomes.

Overall, this evolution of evaluation presents an exciting area for continued research. In many ways these changes in evaluation run parallel with the development of Data Science both as a discipline and as an occupation. Although statisticians and data analysts have been analyzing data to provide actionable insights for a long time, data science represents a formalization of the ways that data analysis methodologies have evolved to more efficiently manage and respond to the large amount of data being generated by new technologies. As I continue my research, I look forward to exploring how the simulation and modeling I have been exposed to through Systems Science and the machine learning and advanced statistical modeling that I have been exposed to through Data Science can be applied to complex problems to have a meaningful impact.
References


ICTR (Institute for Clinical and Translational Research). What are the T0 to T4 Research Classifications?, 2019. URL https://ictr.wisc.edu/what-are-the-t0-to-t4-research-classifications/.


C. K. Prahalad and S. L. Hart. The Fortune at the Bottom of the Pyramid. *Strat-


