Methodological Strategies for Investigating Hard-To-Reach Populations: Essays from Tanzania, Nepal, And Sierra Leone

A Dissertation

SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA

BY

Anna Bolgrien

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Deborah Levison

April 2021

Copyright Page

Acknowledgements

I acknowledge the statistical offices that provided the underlying data making this research possible: Bureau of Statistics, Tanzania, and Nepal Central Bureau of Statistics (IPUMSI data).

I am grateful for valuable feedback and guidance from my advisor Deborah Levison for being an exceptional academic advisor these last many years and a wonderful hiking companion during rare relaxing days off in the field. Audrey Dorélien, Kathryn Grace, Laura Bloomberg, and Greta Friedemann-Sánchez have provided support, guidance, and valuable feedback on this work and mentorship through the entire dissertation process. I'd also like to thank Jennifer Hall-Lande and Sarah Hall for their mentorship through the MNLEND fellowship.

The following sources provided financial support:

Minnesota Population Center
Humphrey School of Public Affairs
University of Minnesota Human Rights Initiative
University of Minnesota Grant-in-Aid
Minnesota Leadership Education in Neurodevelopmental and Related Disabilities
Institute for Social Research and Data Innovation and the IPUMS Center for Data Integration

I am indebted to the field teams in Tanzania and Nepal of the Animating Children's Views project and their dedication to producing high quality data under frequently difficult circumstances such as long days, many hours on the phone, multiple transportation failures, and scary dogs. Tanzania: Julieth Mongi, Gloria Gerald, Haikael Mjema, Rachel Gerald Kishili, Nancy-Nai Laiser, Hellen Urio, Godfrey Rwiza, Tainoi Clemence. Nepal: Sunil Poudel, Pradip Subedi, Sharmila Dhakal, Tejkala Upreti, Kriti Gautam, Rojeena Shrestha. Additionally, thank you Hillary Carter-Ligett, Ebenezer Kweke, Susan James, Jovit Felix, David Mollel, Masai Loreu, Uttam Sharma, and Binu Sharma for their work on the Animating Children's Views Project. Thank you also to Andrea Ferro and the students at UFSCar -Sorocaba in Brazil.

I'd also like to thank the support of the MPC trainees, including Sarah Garcia and Maryia Bakhtsiyarava, and my Humphrey cohort for the joy and conversations that make graduate school infinitely more rewarding.

Personally, I would like to recognize my parents, Kelly and David Bolgrien, for their unconditional support, and Bjorn Gowdy-Jaehnig for his encouragement and endless patience.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Demographers seek to identify, appropriately measure, and accurately count human populations. Data and analyses produced by demographers are essential for researchers working in many disciplines in social science and adjacent fields, generally described as "population science" (Xie 2000; Morgan and Lynch 2001). Objective demographic quantification is essential in order to represent, generalize, and validate the forms and functions of human interactions within and between populations. The movement towards the quantification of more social science phenomena (Espeland and Stevens 2008; Berman and Hirschman 2018) has resulted in demographers being integral to the production of policy-critical data and policy making.

As an interdisciplinary scholar of demography and public policy who uses a variety of quantitative and qualitative methods, I recognize that the co-evolution of objectivity and influence has led some scholars in the field to argue that demography and population studies are constrained by theoretical limitations or dependent on social constructions. For example, Hodgson (1983) argues that demographic statistics are inherently political as they are used to shape policies such as family planning or development initiatives, and Watkins (1993) argues that demographic representations of gender, fertility, and family are shaped by a patriarchal lens of gender roles. Saltelli (2020) suggests that demographic quantification in the era of big data masks both the influence of subjectivity and potential for improper statistical modelling. I have taken pains to recognize and bring in these critiques throughout my dissertation and related research.

The logistics of sampling *hidden* or *hard-to-reach* populations exacerbates the potential limitations of quantitative and demographic methods. In the literature, hard-to-reach populations are broadly defined. In some situations, "hidden" refers to populations that are actively hiding actions or characteristics from data collection due to social pressure or stigmatization, such as drug users or persons living with HIV-AIDS (Watters and Biernacki 1989; Magnani et al. 2005). In other research, the phrase "hard-to-reach populations" refers to marginalized or minority people (Brackertz 2007; Marpsat and Razafindratsima 2010). The challenges of deploying quantitative methods in studies of hard-to-reach populations results in a systemic lack of reliable data about such populations. This is particularly acute in developing countries where large-scale

demographic data are less frequently collected (Jerven 2013). Reliable participation of populations in the collection of demographic data are foundational to subsequent quantitative studies of health, industry, development, and socioeconomics (Teitler, Reichman, and Sprachman 2003; Fitzgerald and Fuller 1982).

In this dissertation, I present three essays on populations in developing countries that can be considered hidden or hard-to-reach: children 12-to-17 years old, siblings of persons with disabilities, and pregnant women during an epidemic. These populations may not be hidden or hard-to-reach in a traditional sense; they are simply accidently overlooked or, insidiously, purposefully ignored by academics and policy makers alike. In each chapter, I highlight reasons these populations have been overlooked by policy makers and document the value of demographic data about these populations that are "hidden in plain sight." I suggest methodologies by which quantification and data collection can improve policy.

In Chapter 2, I explore the social and logistical challenges of including children in survey research in Tanzania and Nepal. While children are not usually considered a hidden population, I (and others) argue that it is their *perspectives* that are hidden, particularly those of children in developing countries (Grover 2004; James 2007; Borgers, De Leeuw, and Hox 2000). Scholars of the New Sociology of Childhood and participatory research propose that the perspectives of children are typically discounted and often ignored entirely by policy makers (Hill et al. 2004; Tisdall and Punch 2012; Kelley 2006; McKendrick 2001). Taft (2019) clarifies that such behaviors are rooted in assumptions that children are fundamentally different from adults and therefore inevitably subordinate. Their lives are assumed to be separate from or excluded from an adult-centric world. Children make up a cohort of individuals who have unique perspectives and experiences. That *should* be given weight on a large-scale, as they are unique to a child's age and the era in which they live (Holt 2006). These perspectives can have profound effects on other things that demographers and policy makers care about such as marriage, work, reproduction, and relationships.

I describe several of the main social and logistical difficulties of sampling children for survey research. The sampling and data collection process described are based on three studies conducted in rural and urban Tanzanian and peri-urban Nepal. The sampling process sought to include children 12 to 17 years of age in a household-style survey comparable to large, nationally-representative surveys. I construct measures of survey quality and data quality typically used by demographers and policy makers to assess the success of the sampling processes: outcome measures such as response rates, success rates, refusal rates, and noncontact

rates, sample weights, and population estimates. The inclusion of children in large-scale demographic surveys allows for quantitative data on children's perspectives to be used by policy makers. This can help children go from being a hidden population to active and agentic members of the policy realm (Holt 2006).

In Chapter 3, I focus on the hidden population of siblings of children with disabilities, using census data from Tanzania. People with disabilities are often called the largest minority group in the world, though rarely studied in developing countries by demographers due to difficulty with cultural interpretation and quantifiable measures (Chamie 1989; Fujiura, Park, and Rutkowski-Kmitta 2005; Fujiura, Rutkowski-Kmitta, and Owen 2010). Even less attention has been paid to the families of persons with disabilities, specifically siblings, in developing countries. This is a problem because it is well established that familial support for persons with disabilities is often their most important form of support in developing countries, given the lack of governmental services (Heather M. Aldersey 2012). The lack of governmental support is one way that policy is marginalizing and making invisible persons with disabilities and, by extension, their families. The hidden and unrecognized support that siblings of persons with disabilities provide, such as caregiving responsibilities, can have additional implications for the lives and experiences of children without disabilities.

In Chapter 3, I use Tanzanian census data to identify and to establish descriptive estimates of the number of children living with siblings with disabilities. In 2012, 3.5 percent of children under the age of 18 either had a disability or had a sibling with a disability (approximately 510,000). I also explore the relationship between having a sibling with a disability and education for children in Tanzania. Specifically, I analyze the probability of being enrolled in school for children who have siblings with disabilities compared to children without siblings with disabilities. The analyses rely on logistic regression to estimate the odds of being enrolled in school across various operationalizations of sibling status. Sibling and family inter-relations are important contributors to many demographic measures but are often overlooked in disability research conducted using large-scale quantitative data. This chapter highlights the importance of considering sibling relationships in developing countries as a way to bring to the attention of policy makers the experiences of persons with disabilities and the families that support them.

The final essay in Chapter 4 approaches the idea of an overlooked or hidden population in a slightly different way (compared to the other chapters in this dissertation). The chapter explores the effects of the 2014-2016 West African Ebola epidemic on the fertility of women in Sierra Leone. The long-term social and demographic effects of Ebola are still being uncovered in

the wake of the medical response to contain and control the outbreak (Farmer 2019). I argue that fertility was overlooked during an epidemic of the size and severity of the Ebola outbreak (Schwartz, Anoko, and Abramowitz 2019).

In this chapter, I use nationally-representative data to show decreases to population level fertility in Sierra Leone and increases in the interval between births for individual women living in Sierra Leone during and after the Ebola epidemic. The analysis in Chapter 4 utilizes a difference-in-differences approach to look at fertility changes across districts before, during, and after the Ebola outbreak. Additionally, an event history analysis explores the social and behaviors choices women made about their own fertility during the outbreak. Policy makers aiming to make effective public health and social medicine responses to epidemics must incorporate a more holistic approach to care in order to prevent harm of populations from being lost, ignored, or hidden.

In each of the three essays of this dissertation, I reflect on methodological improvements for identifying populations hidden in plain sight. I balance the influence of social constructs with objective and well-founded statistical analysis to show that demography remains necessary for policy making despite the social and theoretical influences. The focus on populations in developing countries – specifically Tanzania, Nepal, and Sierra Leone – promotes further policy attention at the intersection of childhood, health, and family. Children, health, and families are at the core of demographic data used to make decisions about policies. If children, siblings, and mothers are systematically overlooked or hidden in demographic data, they are excluded from the very policies that claim to support them. The fact that such populations' perspectives and experiences are hidden in plain sight is a challenge to demographic objectivity. Interdisciplinary scholars, like myself, must establish demographic data that reveals these "hidden" populations in order for them to be seen by policy makers, while recognizing the social and theoretical critiques of demographic objectivity.

Chapter 2

Measures of Sampling Success: Practical and Social Challenges of Designing and Implementing Sampling Frames in Three Case Studies Targeting Children in Tanzania and Nepal

Abstract: The socio-political and logistical practicalities of field work affect how sampling and data collection are conducted, particularly in developing countries. This paper details the sampling processes in three case studies – proportional stratified sampling (rural Tanzania), multi-stage sampling with two rounds of randomization (urban Tanzania), and geographically dispersed sampling with three stages of randomization (peri-urban Nepal). I argue that transparency and specificity about the sampling processes are required in published research, as these processes directly impact other types of reported measures of survey success. Analysis of measures such as success rates, refusal rates, and nonresponse rates in addition to the creation of sample weights and population estimates show the varying complications and reality of each sampling strategy. I advocate a more stringent level of transparency when reporting about survey measures in academic publication.

2.1 Introduction

Calls for greater transparency in research practices and online publication of raw data are transforming research in all fields (Schooler 2014; Grahe 2018). Survey methods are also changing; moving from pen-and-paper surveys to digit tablets or online platforms allows for more data *about the survey* to be collected, stored, and analyzed (Hughes, Haddaway, and Zhou 2016). This data about the survey is categorized as *metadata* (data about the data) and *paradata* (data about the process of collecting the data). The accessibility of metadata and paradata allows researchers to validate the data collection process with greater integrity (Couper 2005). Utilizing paradata and metadata in sampling and survey data collection is one answer to the call for greater transparency, but it is not a panacea as processes of data collection and sampling are still

marginalized in most social science literatures and practical strategies and best practices are rarely taught in classrooms.

Survey technology, with the ability to utilize metadata and paradata is now affordable and accessible to even small-scale researchers. But there are not clear guidelines on what to do with paradata (Lynn and Nicolaas 2010) The theoretical recommendations that do exist are often targeted at large-scale, nationally representative surveys. Detailed sampling frames and methodological reports published about these large projects outline cross-country comparability and high-level theoretical equations applied to large sample sizes[1]. But most researchers, particularly graduate students or early-career faculty, will not be involved in data collection on a large scale because of limitations of cost, budget, time, collaborators, expertise, or niche topics. Individuals aiming at independent field work and data collection struggle to apply literature and theory describing large-scale surveys to answer practical decision-making concerns in small-scope projects.

Small pilot projects are the bedrock of academic curiosity and exploration but lack the high-profile status of well validated and published large-scale quantitative research. Researchers learn many practical aspects of designing and implanting sampling frameworks via small-scale studies. New theories or hypothesis can be tested to provide justification and support when applying for grants or planning larger studies. Small survey projects also accompany qualitative work and mixed methods work (Onwuegbuzie 2007). Practitioners often learn from doing as they gain on-the-ground experience with the messy reality of fieldwork and adapt sampling methods to the local socio-political environment. Yet when it comes to publishing academic work, the messiness of the sampling process is often sanitized through the careful reporting of handpicked success rates that smooth over the details and complexity of the sampling processes. To discuss vulnerabilities in the data is to risk rejection by journal referees. As a result, first-time survey researchers doing small projects lack published examples. This creates a gap in the literature right at the intersection of validation and transparency.

This paper describes three small-scale pilot studies and highlights challenges of field research by taking a detailed look at the entire process of collecting survey data: from the decisions over what type of sampling strategy to use, through the implementation of the strategy in the field, and finally to the reporting of successes and failures of that process. This holistic view is aimed at

[1] Some examples of cross country surveys conducted in developing countries include the Demographic and Health Surveys (The DHS Program 2021), the UNICEF MICS surveys (UNICEF 2020), and the PMA2020 surveys (Zimmerman 2017)

development practitioners working at intergovernmental or nongovernmental organizations or independent researchers who intend to publish in academic journals.

Through three case studies, this paper describes the sampling processes used in the Animating Children's Views (ACV) project in rural and urban Tanzania in 2018 and peri-urban Nepal in 2019. The project's goal was to produce random sample of households in each of the case study locations[2]. However, due to the physical logistical limitations and social structures of geo-political life in each location, this proved to be complicated. Building on traditional methodologies for sampling frameworks in developing countries, the three case studies implemented three different sampling strategies: proportional stratified sampling (rural Tanzania), multi-stage sampling with two rounds of randomization (urban Tanzania), and geographically dispersed sampling with three stages of randomization (peri-urban Nepal). This paper will describe the logistical processes of sampling in each of these three case studies and consider how differences in social structure affects decisions related to the sampling and data collection (Parts 2.2-2.4). The results of the sampling process are summarized according to the Total Survey Error framework (Part 2.5). Then I compare the results of data collection for each of the cases studies through various reporting measurements common in academic publications, such as success rates, response rates, refusal rates (Part 2.6) and produce sample weights and population estimates (Part 2.7-2.8) constructed using survey paradata. By constructing these measures, I evaluate the goal of creating comparable samples in three sites, though we do not claim to generalize to a broader target population nor intend to pool the data across the pilots. I suggest the measures described – success rates, response rates, refusal rates, sample weights, and population estimates – are commonly oversimplified in published research in ways that mask the complexity of the data and data collection process. Finally, I propose guidelines that indicate which measures allow for transparency in cases of sampling designs for small scale projects (Part 2.9).

2.1.1 Total Survey Error Framework

The field of survey research aims to improve sampling and data collection methodology and decrease errors in statistical measures. The Total Survey Error (TSE) framework established by Groves (Groves 2011; Groves and Lyberg 2010) provides a theoretical understanding of the

[2] At each household, one adult (usually the mother) was surveyed in addition to all children age 12-17-years-old who resided in the household. This paper describes only the process of sampling and surveying households, not specifically the individuals within the household. Unless otherwise specified, a sampled and surveyed household refers to a household where the field team interviewed one adult and at least one child in the age range.

process of conducting a survey and identifies areas of potential bias [Figure 2.1]. Briefly, conducting a survey has two major potential components: measurement and representation. The measurement side encompasses the validity of the survey instrument in capturing data that accurately represents the concepts it intends to measure. The representation side include the progression from the target population to the sampling frame, sample, and respondents. One major critique of the TSE is the lack of quantifiable measurement recommendations for the different types of error found along the process of conducting a survey (Groves and Lyberg 2010). Nonetheless, it can be a useful tool for survey practitioners to explicitly explore the possible errors in their survey.

In this paper, I will use the TSE framework to conceptualize areas of potential bias and error in the ACV pilot studies specifically on the representation side. I will not address issues of measurement and instrument design. Within the representation side of the TSE, there are three main sources of potential error:

- *Coverage error* occurs when establishing a sampling frame from a target population
- *Sampling error* occurs when producing the sample from the sampling frame
- *Nonresponse error* occurs when identifying and collecting data on respondents from the sample.

I will first discuss coverage and sampling error in reference to the sampling processes of three ACV pilot studies in the following sections. Nonresponse error will be addressed in greater detail in section 5 as I discuss different potential challenges of identifying, finding, and surveying respondents from the sample.

2.1.2 Background on Sampling Frames

A sample, in survey research, represents a population without conducting a census (Hubbard et al. 2016). It is important to first establish some general terminology. The *target population* is the population of interest for the survey. For example, the target population could be all households in a specific geographic area or all individuals with a shared characteristic, such as attending a specific school. In the ACV pilot, the target population is all households with at least one child age 12-to-17 living in specified geographic areas: a village in rural Tanzania, a specific urban area in Tanzania, and selected municipalities in Nepal. A *sampling frame* identifies all eligible units (i.e. households, individuals) of the target population. Ideally, this is a complete list of all members of the target population; for example, a list of all households and

household members in the geographic area or a full roster of all students attending a school. The sampling frame is rarely a complete list and is often constructed via probabilistic selection or from multiple sources. From the sampling frame, a *sample* is drawn of individual units (households, individuals) who will be contacted to participate in the study. The sample is a representation of the target population drawn from the sampling frame using a probabilistic method[3].

When a sampling frame contains complete information about the target population, a simple random sample can be drawn from the sampling frame. However, this process is rarely as easy or as simple as it may seem. Random sampling in developing countries can be difficult, depending on availability of current population data, accurate spatial boundaries, and clearly organized and labelled households and communities. In areas that lack accurate or complete sampling frames, additional sampling techniques must be applied to create a representative and probabilistic sample of the target population. At the forefront of the research on developing sampling frames are epidemiologists, who generally use the Expanded Program for Immunization (EPI) framework developed by the World Health Organization in health studies; an early example of this method found in (Henderson et al. 1973). EPI sampling methods generally have a two-stage (or more) sampling process. First, communities or clusters are purposefully or randomly selected within a larger geographic area. For example, a sample of villages may be selected within an entire country. Following the identification of communities, the EPI method requires either a complete list of households in the target population – which can be all households or households with a specific sub-population characteristic (e.g. children 0-5) – to create a random sample or another method of randomization can be used to determine which households to sample. Few communities have a such a list readily available; often, the only reasonable alternative is to conduct the census of households oneself, which can be expensive and time consuming. If this cannot be done, sampling methods such as a random walk or "spin the pen" method may be used to identify households, though these techniques are subject to criticism of their possible lack of probabilistic nature (Grais, Rose, and Guthmann 2007; Bauer 2016). Another option is to work with a national statistical office, which often requires complicated social relationships and a recent national census.

The EPI sampling method is considered the standard for sampling and has been adapted by many major cross country survey organizations in developing countries, including the

[3] Summarized from Survey Research Center (2016).

Demographic and Health Survey Program (The DHS Program 2021). Modifications of this method have generated a rich diversity of sampling frames reminiscent of EPI as researchers adapt the method for the inclusion of new technologies such as GPS and satellite imaging (Haenssgen 2015; Wampler, Rediske, and Molla 2013; Kondo et al. 2014) or greater statistical specificity (Turner, Magnani, and Shuaib 1996; Milligan, Njie, and Bennett 2004). New technologies such as computer-aided personal interviewing (CAPI) have allowed researchers additional tools for sampling, tracking data collection, and verifying data quality (Savel et al. 2014; Hughes, Haddaway, and Zhou 2016; Caviglia-Harris et al. 2012; Abelsæth 2012).

While the use of technology moves the field of sampling and survey research forward, these modifications to accommodate more complex probabilistic sampling are challenged by the reality of conducting research in developing countries. The demands of each location are accompanied by limitations including lack of accurate and up-to-date data from government officials and cooperation or resistance of local leaders. Large data projects also tend to hire large field teams; it is difficult to completely account for differences among individuals doing the data collection and sampling, despite best efforts of training and streamlining survey procedures. The social and human element plays an important but underreported role in the success or failure of any field work project.

2.1.3 Animating Children's Views Project

The goal of the three pilots described in this paper was to produce three studies that were of similar size and deployed similar research designs. The Animating Children's Views (ACV) project is a mixed methods study that developed a new survey methodology; it uses cartoon videos to survey children about their views and perspectives on issues that are facing young people. The ACV pilots establish a methodology that could be expanded to a national or cross-national scale (Levison and Bolgrien 2020). The methodology is currently being tested in small pilot studies designed as household surveys, for eventual use by large-scale survey operations. Person-to-person interviews use the tablet-based survey software SurveyToGo (Dooblo, n.d.). Built into the SurveyToGo software are quality check measures that track time spent on each question, possible modifications to answers, or falsification of data. Although the ACV project is small in scope and overall budget, the project sought to mimic a large representative household survey in both design and sampling strategy using EPI and other sampling literature as the foundation for developing context specific sampling frames. The project sampled from the target population of households with 12-17-year-old members. These pilots provide realistic examples

from which to examine and critique the process of applying textbook strategies in a complex and messy world, collecting data with small teams on limited budgets in three very different places.[4]

Many excellent textbooks and review articles have outlined different kinds of sampling frameworks (Johnson et al. 2019; Kish 1965; G. Kalton 1983). To use the language of Fottrell and Byaas (2008), the rural Tanzania pilot used a proportional stratified sample, the urban Tanzania pilot used a multi-stage sample where both stages included randomization, and the peri-urban Nepal pilot used a geographically dispersed sample with two stages of randomization and a random walk. For reference, the stages of sampling are outlined in Table 2.1. The literature relevant to each of these strategies will be outlined in more detail below. Each of these methods establishes a strategy that creates a sampling frame from a target population and then conducts a sample from the sampling frame.

2.2 Pilot 1- Rural Tanzania

Tanzania was selected as the first country to pilot the new ACV methodology. The East African country has a strong system of bureaucracy and local leadership within small communities of people. There are 30 major regions, divided into 169 districts that are divided into municipalities. In rural municipalities, villages are further divided into sub-villages. Urban municipalities are divided into wards, then "streets" (called *mtaa* (singular) or *mitaa* (plural) in Swahili), and finally ten-cells (originally groups of 10 households). Theoretically, there is a "ten-cell leader" who is responsible for knowing the identities of ten households living within a small area like a block; however, the size of cells varies greatly depending on the urban areas. The sub-village and ten-cell leaders are responsible for keeping lists of people in their cell or sub-village. Thus, Tanzania appears to be an ideal context to use a sampling strategy that relies on accurate household lists of current populations living in small spatial units even when national census data is out of date or not at a small enough spatial geography. The last census in Tanzania was conducted in 2012; since then the country has experienced a lot of population growth and migration, so the 2012 population figures were not likely to be accurate in 2018.

[4] This project is part of an ongoing protocol, approved by the Institutional Review Board at the University of Minnesota. The research was also reviewed for ethical and social appropriateness by COSTECH (Commission for Science and Technology) in Tanzania, district and municipality offices in Tanzania, and municipality offices in Nepal. Oral informed consent was obtained from local community leaders as they assisted in the sampling process in Tanzania.

In Tanzania, the research team worked with a local survey research organization to hire a team of field researchers to conduct the sampling and the data collection. ACV conducted two pilot studies in Tanzania: one in a rural village and one in an urban city, both purposefully selected within Arusha Region in Northern Tanzania. The aim of these pilot studies was for the ACV study to mimic a large-nationally representative survey while being limited by a realistic budget and time constraints. We wanted to identify pilot areas diverse in ethnic groups, religions, and livelihoods located within the study area. By conducting pilots in the same region, we spent less time obtaining approvals and permission letters from regional officials. The target population of the ACV project is households with children 12 to 17 years of age. The intention was to survey one adult household member, preferable the mother, and at least one child in the 12-17 age range. We used probabilistic sampling; however, as described below, the project can only generalize within the geographic areas we worked in and not for Tanzania as a country. The PI (Prof. Deborah Levison) and I were onsite to supervise the data collection process, with daily debriefings.

2.2.1 Village Selection in Rural Tanzania

In the rural Tanzania ACV pilot, we purposefully selected a single village for data collection. The village selected is adjacent to a small urban center along a major highway[5]. The village was selected based on previous knowledge of its population as being diverse in religions, tribes, and livelihoods (Ritter et al. 2010). It was also selected for logistical reasons: it was within a single day's drive from a major city where the team was based, the road to the village was passable in July, and the area had cell service to allow the team to communicate. Sampling and data collection were limited to a single 12-day time frame because of time and budget restrictions.

2.2.2 Sampling Process in Rural Tanzania

We used a proportional stratified sample based on the structure of rural villages in Tanzania. This strategy divides the village proportionally to the population of its sub-villages. The population of sub-villages varies; thus, a simple random sample would result in higher-population sub-villages having a higher probability of their households being selected. Proportional stratified sampling maintains the proportional share of participants by mandating a proportional number of households in each sub-village, the *stratum*, be included in the final

[5] The village is unnamed for confidentiality reasons

sample. Randomization occurred within each of the sub-villages. Figure 2.2 show the process of sampling in rural Tanzania.

Each of the seven sub-villages in the village selected for the pilot was represented by a sub-village leader. After confirming cooperation from village leaders, we asked each sub-village leader to provide the research team with a written list of the persons and households living in their sub-villages. If a list did not exist already, we paid sub-village leaders for their time and help preparing the lists and identifying households. This process followed standard recommendations for EPI sampling to conduct complete enumeration of a selected area (Milligan, Njie, and Bennett 2004; Turner, Magnani, and Shuaib 1996). The success of this process will be discussed in further detail below.

For each sub-village, we counted the number of households and people identified by the sub-village leader. To conduct our study, we needed to sample only among households with children ages 12-17 years old. To do this, we assigned each household on the list (those with and those without children ages 12-17) a number. Then, using a random number generator, we identified households based on the order of the random number generator, keeping only those households with children ages 12-17 on our final list.

To implement the proportional stratified sample, the number of households per sub-village was capped based on a proportion of the total population of the village. These proportions were used as guidelines for the number of households successfully interviewed in each sub-village. The aim of our study was to survey approximately 100 households with children total in the village. On average, we needed 17 households per sub-village with smaller sub-villages needing a minimum of 13 and the larger sub-villages needing a maximum of 25 to be proportional to the overall population size of the village. We complied lists of between 20-30 total households with children in each sub-village in order to account for refusals, inaccurate reporting of children's ages on the household lists[6], and other problems such as not being able to find the physical location of the households or not being able to find the members of the household. In some sub-villages, listing 20-30 households with children 12-17 ended up being almost a census of households with 12-17-year-old children due to smaller overall population

[6] In the process of creating the sample frame from the sub-village lists, we knew that there would be households identified as being in the target population of those having a 12-17-year-old but that did not actually have a child of that age. It happened occasionally that the field team would arrive at a household and find that the intended child was actually 10, 11 or 18 years old. As we were able to anticipate this in advance, we were able to construct our sampling frame to accommodate this situation to reduce the potential of over-coverage affecting our coverage error.

numbers. For all sub-villages, we achieved the desired number of households as proportional to the population of the village.

2.2.3 Social and Logistical Challenges in Rural Tanzania Sampling and Data Collection

Requesting and maintaining the assistance of local sub-village leaders during the sampling process was the first of many social challenges we faced as we conducted the sample and collected the data. The first step of creating a sampling list is to create an accurate sampling frame. Ideally to reduce coverage error, we hoped to create a sampling frame across the entire village before starting the sampling and data collection in any sub-village. We relied on the knowledge of the sub-village leaders to provide us with accurate information about the population. Some sub-village leaders were prepared and willing to share their lists openly with the research team. Others didn't have lists and took several days to go door-to-door to enumerate the households. One leader would bring a few handwritten pages of lists on one day and then the next day, he would bring a few more. This resulted in lists being created while data collection in other sub-villages had already started. It was never clear if we had a complete list from this sub-village as the patience of the sub-village leader waned as the days went on and his enthusiasm for helping us diminished. There were also published figures of the population posted in the village office that provided a finer level of detail than figures from the last population census in 2012. We were not able to confirm exactly the date of publication for these posted figures as they varied substantially from the 2012 census numbers and from lists we collected from the sub-village leaders. These concerns fall into the Total Survey Error (TSE) framework potential for coverage error as they pertain to our ability to create a sampling frame from the target population.

In several situations, we completed the sampling frame but were missing several other pieces of information that would assist with the creation of sample weights and outcome responses during the analysis. Partial information varied across sub-villages. For example, for some sub-villages we recorded only the total number of households with children 12-17-year-old members but are missing information regarding the total number of households, and vice versa in others. Some of the sub-village leaders were only available to assist us on specific days, and we were unable to ask about the missing information or confirm our numbers within the data collection time frame. When we did not have proper enumeration of the target population results, overall population figures are estimated based on the information that was gathered when computing outcome measures such as response rates and survey metrics such as weights (discussed in detail in sections below). The social interactions needed to access this information,

as described in the process of creating and generating the household lists, greatly depended on the sub-village leaders' interest and availability in working with the research team. The population figures and full enumeration counts used to establish the proportional number of households needed in each sub-village resulted in underestimating the total number of sampled households we expected to be able to find in some sub-villages and overestimating in others.

In the TSE framework, sampling error may occur in the process of establishing the sample from the sampling frame. In addition to social limitations, during the sampling process there were other practical and logistical challenges. First, though our aim was to find children ages 12-17-year-old, we were limited by scheduling conflicts such as school hours and extra tutoring sessions attended by many of the children. It was important for the success of the project to interview the children in a non-school setting. We had to work around the school schedule to find times when school-going children were at home. Additionally, when we were selecting the village, we were not aware of a policy that required all secondary school children in the municipality to attend boarding school. This is discussed in greater detail in sections below. Thus, the sample in the village pilot is missing many older children who were away from home and could not be interviewed in the 2-week period we were in the village. The limitation of school children and boarding school attendance was a logistical challenge in the survey process that potentially affects the sampling error.

Another logistical challenge arose due to transportation and budget issues. As we paid sub-village leaders to assist the data collection team in finding households, travelling between locations was time intensive and costly. Houses in the village were often spread out. The survey team was small and had only one vehicle. This also resulted in a lot of time that some members of the team spent waiting for others to finish interviews.

Concerns about safety also limited the team's activities. In order to keep the team safe, we encouraged all interviews to conclude by sunset, which was approximately 6:15 pm in July. This protected team members but also severely limited the time interviews could be conducted with children after they returned home from school and before they were expected to do chores and other work responsibilities at home. Most of the interviews with adults happened during the day while children were at school. This allowed the field team to prioritize the interviews with children outside of school hours, but it created the additional barrier of needing to return to a house multiple times to meet with the adult and then again with the child or children. Many households also lacked electricity, which made it difficult for the field team to do their job and

also may have created an uncomfortable environment for children being interviewed by strangers at dusk or, occasionally, in the dark.

In this first case study in rural Tanzania, the success of the survey team to create a proportional-stratified sample and carry out quality data collection required adapting every step of the process to the local village context. The ability to use the sub-village enumerated lists greatly helped simplify the creation of the sampling frame. This approach was only possible based on the bureaucracy of the sub-village system in rural Tanzania villages. However, even with this seemingly straightforward approach, the social and logistical challenges of field work shaped the data collection process and any potential coverage and sampling error. The quality of the data, metadata, and paradata was entirely dependent on the dynamics between the field team, the sub-village leaders, and the respondents working together to make the sampling and data collection a success.

2.3 Pilot 2 – Urban Tanzania

2.3.1 Selection of the Urban Tanzanian Pilot Location

The second pilot was in the booming northern city of Arusha (population: 416,442[7]) that is growing rapidly as people migrate from rural areas. As in the village, the target population was households with at least one 12-17-year-old resident. Urban cities in Tanzania also operate within a hierarchical political system. This benefited our data collection, as it was not possible for us to sample the entire city on our budget. Unlike in the village pilot, it was not possible to conduct an enumeration of the entire city of Arusha. The first task of the ACV pilot was to identify a sampling unit that was small enough to have the household list that could be used as a sampling frame. Areas of Arusha were selected though a multi-stage sampling with two stages of randomization. This sampling method, like the multi-stage sampling technique in Fottrel and Byass (2008), is similar to the original EPI method of sampling which calls for the identification of clusters or strata and then the complete enumeration of the clusters in order to produce a household list; an example of this process is also described in Alves et al. (2012) in Brazil.

[7] From the 2012 census publications

2.3.2 Sampling Process in Urban Tanzanian Pilot

First, it was determined that we would only sample within a purposefully selected municipality that represented the urban areas of the city. Like the rural village, the results from the urban Tanzania pilot can only be generalized to Arusha. There were many challenges in determining the study site because district and municipality lines do not exactly match with other geographic political borders, depending on the source. For example, the city of Arusha is located in Arusha Region, but we had to compare specific wards in order to determine which geographic area to sample: Arusha Municipality, Arusha Urban Municipality, or Arusha City. These three names represent similar, but not identical, geographic areas depending on political units that were not always clear to outside researchers[8]. Finding accurate lists of geographic areas required a local collaborator to make many trips to the municipality office. The sampling frame was based on the geographic boundaries of Arusha City (Arusha Mjini).

Within an urban municipality, the next geographic unit is the ward. Across the 19 wards in Arusha, there are 125 *mitaa* (singular *mtaa* in Swahili and generally translated as "street"). Within each of these *mtaa*, the smallest unit of geography is the cell; in an urban area the cells vary between having 10, 50, or even 100 households with one politically appointed cell leader. Figure 2.3 shows the sampling process in urban Tanzania.

We randomly selected 23 *mitaa* to visit. Two *mitaa* were excluded for being too rural and one was specifically used for training purposes. This left 20 *mitaa* in our sample. In each *mtaa*, local collaborators asked the *mtaa* leader to make a list of all of the cells and cell leaders in the *mtaa*. Then, we randomly selected one cell within the *mtaa*. Within this identified cell, we requested that the cell leader create a list of households with 12-17-year-old children in the cell. This resulted in 22 eligible households on average [min:7, max: 58] and we randomly selected 10 to be included in the sample. We allowed substitution of additional households if fewer than 6/10 of the originally selected households were found to be eligible[9]. Through this multi-stage process, randomization occurred at the *mitaa*, cell, and household level.

To make these results comparable with the pilot in rural Tanzania, we sought to interview about 100 households with children across the city. The sampling distribution was designed to be equal across each of the *mitaa*, instead of proportional like in the village study. We hired several

[8] This is similar to the differences between counties, school districts, and congressional districts in the USA. It is imperative that the exact boundaries and units are known before proceeding with a sampling frame.

[9] If this did not lead to enough eligible households, we would have moved onto the next identified cell. But this step did not occur.

local collaborators to go in advance to identify households and collect contact information prior to the start of data collection. We attempted to identify households willing to participate before sending the team, in an attempt to save the research team's time. This resulted in high participation rates though it was time intensive and costly. We paid the cell leader to take us to residents' homes and make introductions with sampled participants.

2.3.3 Social and Logistical Challenges in the Urban Tanzanian Sampling and Data Collection

This multi-stage process was time consuming and required multiple visits to different areas of the cities to allow cell leaders to create the households lists that were sampled from. While we have trust in our local collaborators, we don't have a good understanding about the quality and completeness of the lists produced by the *mitaa* leaders and/or cell leaders. The city is expanding so rapidly due to migration that it is difficult to expect the leaders to know their neighbors in the same way they can in a village or in a smaller urban area. Both of these issues could potentially increase coverage error in our pilot.

Even within one country that relies on the same hierarchical political structure, the process of sampling in rural and urban Tanzania required different sampling strategies. Social challenges of working with local leaders included travelling between *mitaa,* visiting offices, rescheduling appointments, explaining and reexplaining the purpose of the study, and following social norms of respect. All of these steps were necessary, but they required a significant amount of time, energy, and money before conducting a single interview with respondents. Every step influenced the success of data collection and affected the potential for error in the survey statistics. For example, we did not track metadata that recorded the number of attempts to contact each household or the number of visits to each household. It is possible that multiple visits to a household may impact the sampling error on the pilot based on unidentified differences between households that were available for surveying on the first attempted visit, households that were eventually identified after multiple visits, and households that were never identified.

Similar to the village pilot, in the city we faced logistical challenges of transportation, daylight hours, and safety described above. Field team members carried relatively expensive tablet computers, which led us to pay for private taxis instead of asking the team to use inexpensive, but somewhat erratic, public buses. In order to maintain social relationships with

local cell leaders and respondents, it was important the field team arrived to scheduled meetings on time.

Respondents in the city tended to be busier and away from home for longer hours than respondents in the village. The team worked hard to accommodate schedules including school hours for the children. Unlike in the village pilot in Tanzania, the urban pilot was scheduled during a vacation time for many students. This benefited the team as it was easier to find children at home during the day, including secondary school students who had returned home from boarding schools for the break. Finally, one benefit of constructing a sampling frame in advance of data collection was that the field team had personal phone numbers for the respondents, given with permission in pre-data collection contact and consent processes. This allowed the team to call respondents in advance.

2.4 Pilot 3 – Peri-urban Nepal

Nepal was selected as the second country to be included in the ACV pilot projects. The success of the project in Tanzania needed to be replicated in a completely different context in order to show the viability of the ACV project globally. The cultural and social traditions of Nepal vary greatly from Tanzania while still providing a context where many children face difficulties in day-to-day lives. We selected Kathmandu, the country's capital and largest city, as the focus for our study.

2.4.1 Selection of Peri-Urban Pilot Location in Nepal

As the second Tanzanian pilot tested the ACV methodology in an urban setting, the third pilot aimed to identify a peri-urban or suburban area of Kathmandu. These peri-urban areas are home to a mix of people, both new migrants and multi-generational residents. The city is expanding into the hillsides of the Kathmandu Valley, and areas of jungle and rural villages are now booming with construction and people. The very shape of the Kathmandu Valley is conducive to peri-urban settlements. The central city of Kathmandu is enclosed by a circular road system, called the Ring Road, with road "spokes" that extend into the hillsides and into surrounding municipalities. In Kathmandu District, which includes Kathmandu City, there are 11 total municipalities. Our target population in the third pilot was households with children 12-17 years old within two purposefully selected municipalities outside of the Ring Road.

2.4.2 Sampling Process in Peri-Urban Nepal

While both pilots in Tanzania had a similar strategy of sampling from household lists, no such political organization exists in Nepal. Without these types of organization, we elected to use a geographically dispersed sampling frame with two levels of randomization. Geographic and spatial sampling has been building on the foundations of EPI sampling as GPS technologies and satellite imagery improve. Researchers can make a sampling frame using GPS and satellites by accessing open source remote sensing data on platforms such as Google Earth or OpenStreetMap. Typically, these maps are used to identify political boundaries and then identify random latitude and longitude within a given boundary in which to start the sampling process on the ground. Alternatively, specific units based on images, such as buildings or plots of land, can be identified to make a sampling list, and randomization happens among these units (Grais, Rose, and Guthmann 2007; Haenssgen 2015). This type of modification to a standard EPI framework has been used in a variety of formats across the world including Malawi (Escamilla et al. 2014), India (Kumar 2007; Montana et al. 2016), Iraq (Galway et al. 2012), Guatemala (Kondo et al. 2014), Lebanon (Shannon et al. 2012), and Haiti (McNairy et al 2019; Wampler, Rediske, and Molla 2013).

Because of budget constraints, we purposefully selected two municipalities with diversity in religions, migrant status, and overall variety in living standards based on recommendations from local collaborators. These municipalities are bounded on one side by the main Ring Road, but they extend far up the hillsides; areas that were formally rural are rapidly becoming peri-urban as the city expands. Once we selected two municipalities, we randomly sampled 50 percent of the wards within each municipality. One municipality had 10 wards and the other had 11; we selected 5 within each, randomly. Figure 2.4 shows the sampling process in Nepal.

I downloaded ward boundaries files from the Nepal government municipal offices[10] and digitized them against existing municipality boundaries. Within the boundaries of the randomly selected wards, all buildings or structures were manually identified, and a list of the building latitude and longitudes was created using ArcGIS and satellite images from OpenStreetMap (Esri 2019; OpenStreetMap Contributors 2019). This process of identifying buildings is preferred to strategies that use randomly generated points within an area. Selection of random points can bias the sample weights as there is a potentially unlimited number of possible points to be selected

[10] Downloaded January 31, 2019 from the Nepal Federal Government Ministry of Federal Affairs and General Administration (copyright 2017).

within a geographic area; thus, it is difficult to tell the probability with which points might overlap actual selection of households (Grais, Rose, and Guthmann 2007). The selection of building structures increased the probability that the point selected would be a residence instead of a location in the middle of a forest or a field and decreased potential coverage error on the sampling frame. The identification of buildings also allowed for the construction of sample weights to control for varying population density across the wards by identifying all buildings and weighting based on additional information gathered in the sampling and data collection process (discussed more in later sections). Once all of the buildings in a ward were identified through this manual process, they were sorted into a random order to provide a basis for the sampling framework.

Buildings identified from satellite images provide a limited amount of information as they often just show the roof of a structure as a two-dimensional rectangle. It can't be determined if the identified structure is a residential building with people living there or another type of structure. Referring back to the TSE framework (Figure 2.1), the target population of households with 12-17-year-old residents make up *some* of the residents in the identified buildings, however a significant amount of work in order to identify the sampling frame and sort out ineligible households. In order to identify if the sampled buildings are residential buildings that also contain anyone age 12-17, the sampler enquired about all eligible and ineligible households in each visited building. This work to identify which buildings contained eligible households is a process where there is a potential for coverage error. In order to minimize coverage error and accurately produce a sample frame, and subsequently a sample, we used a second stage walk from the sampled building (Bennett et al. 1991). Using the list of sampled buildings, a team of samplers went to each of the identified structures based on their latitude and longitude and knocked on doors to see if the building was a residence and if there were any 12-17-year-old children living at that residence. The samplers had a protocol for identifying which building or structure was closest to the latitude and longitude; buildings identified from OpenStreetMaps contained few street names and no building numbers or addresses[11]. Samplers identified all of the households in that building and recorded the presence of children in the households. If there were no children residing at the identified households within the sampled building, the samplers were instructed to follow a protocol to identify other buildings close to the point until a predetermined number of eligible households within the vicinity had been identified.

[11] Nor were street names or building numbers typically available in neighborhoods across Kathmandu.

2.4.3 Social and Logistical Challenges in Peri-Urban Nepal Sampling and Data Collection

In Tanzania, our teams faced social challenges of building and maintaining relationships with local leaders. However, leaders provided the team with legitimacy when interacting with households. In contrast, the field team in Nepal was not introduced to households by local government officials. The samplers and field team had to work hard to establish relationships with each family. This required many phone calls to schedule appointments, early mornings and late evenings to account for working schedules, and long distances traveled to meet with families in person.

There were logistic challenges that came from using GPS coordinates to identify buildings. First, finding the sampled buildings based on a latitude and longitude from a two-dimensional map is very different from finding a three-dimensional building in a physical location. Identifying households required a great deal of perseverance. This was done without any formal street address. Samplers relied on mobile GPS apps like Google Maps to guide them to the specified building. Often buildings were not located on easy-to-access roads or were even located in private gated communities where the sampler was denied entrance. Finding the physical location of a building required samplers to have a strong sense of direction when Google Maps was misleading. These physical limitations took time. For example, a sampler reported a case where a Google Map showed a route to a building that took over 45 minutes of walking to reach, but upon arrival, the sampler realized there was a shortcut that would have taken 10 minutes. As the samplers did not have any information about the household, asking for directions at local businesses or other community members was not helpful, according to debriefing with the sampling team. If a household was not home or available, the team member sometimes relied on neighbors' knowledge of the household in question, particularly if the neighbors resided in the same building. It is uncertain how accurate this information was in some areas.

Once a sampler identified a building at the given latitude and longitude, it was often the case that the residents were not at home at the time, or the building was actually home to several households. Both of these concerns are related to sampling error. Samplers were instructed to try to identify if there were children in any of the families within a sampled building. They were supposed to ask neighbors about any households that were not home at the time. This strategy yielded accurate results in communities there were older, more established, or rural. But because in-migration to Kathmandu is increasing the number of renters, it was often the case that neighbors didn't know the other people living in the same building. Properly identifying

households with children to be included in the final list of sampled households took time and some ingenuity. There was not enough time or money to send the samplers to a location many times.

2.4.4 Validity of Geospatial Sampling in Nepal

One major question affecting the validity of the Nepal sampling process is whether the samplers followed the protocol for identifying households. The protocol established guidelines for samplers to follow after identifying the originally sampled building from the given latitude and longitude. Our random walk protocol stated that the sampler should identify the building at the latitude and longitude given, and then move to the building to the right until the stated number of eligible households are identified. If there was no obvious building to the right (or the building was a non-residential building or open field), the sampler was instructed to move to the left. The intention was the sampler would not stray too far from the original point but also move in a systematic fashion.

We asked each sampler to keep detailed records of the buildings he visited, the number of households in each building, and information about the number of people living in each household he was able to interact with, including eligible and non-eligible households[12]. All of these visited households were reported at the cluster level, as identified based on the sampled GPS point. At each sampling point (i.e. the building identified by the latitude and longitude), we asked the sampler to identify approximately three to five eligible households while staying "within a reasonable distance from the originally sampled point". In some locations, this required the sampler to identify more than five buildings. In other locations, each building contained multiple families and required visiting a fewer number of buildings. Thus, each sampled point resulted in a cluster of households

This process used in the ACV pilot in Nepal varies from the non-probabilistic method of second stage sampling commonly known as a "random walk" or "spin-the-pen" methodology (Grais, Rose, and Guthmann 2007; Bauer 2016). In random walk sampling, the direction in which the sampler would turn after reaching a determined location is randomly decided. In the ACV process described above, the sampler was asked to turn in a systematic direction (to the right) unless he determined this was no longer a useful approach to finding households. The systematic protocol threatens the validity of the "random" walk as the sampler made decisions about which

[12] We hired two samplers who both happened to be male.

buildings to approach or skip in a way that could introduce bias (Chen et al. 2018). For example, if the sampler turned to the left instead of the right when the building on the left appeared to be better kept, we may have too few lower income families in our sample. Because the process of identifying points was done without any assumption that the building identified was a private residence, we also do not have a full picture of how often the originally-identified building was a nonresidential building such as a shop or a school[13]. The exact route of the walked sampling path was not tracked, so we do not know how far the samplers walked from the original point in order to identify the number of households requested. We have to rely on debriefing conversations with samples to assume that the protocol for identifying buildings were followed appropriately and any deviations were either the result of necessity (a dog or security guard prevented access to a household) or lack of a private residence.

Analysis Validity of Nepal Geospatial Sampling

In order to further explore the validity of the sampler's route, I leverage additional geographic information from the survey. The SurveyToGo software used to conduct interviews allows for GPS capture at the site of the survey. Most of the interviews were conducted in the home of the family. I compare the GPS location of the home of the respondent to the originally sampled point (building). This analysis provides a measure of the distance between the originally sampling point and multiple surveys conducted in different households that were identified based on the sampler's walk protocol.

Using the Generate Near Tool in ArcGIS, I identified the sampled point closest to each survey location. In 76 percent (n=118) of the interviews, the survey GPS point was paired with the expected sampled GPS point. Among these points, the average distance (as the crow flies) from the surveyed households to the point is 56.5 meters (min = 2 meters, max=307 meters). This indicates that the sampler's walk remained close to the sampled point and the interviewed households were associated with an area near the building.

The household surveys that were not paired with the expected GPS point fall into three possible categories. First, some households (5 percent, n=9) were interviewed in a different ward than the anticipated ward based on the sampling list. Most of these respondents, based on documentation from debriefing, are assumed to have been interviewed at a location different from their residences. For example, a respondent could request to meet at her place of work or a local café or community center. Another possibility for variation in the expected wards would be the

[13] Buildings labelled as nonresidential were skipped when tagging buildings in OpenStreetMaps

cellular reception of the tablet computers used in the survey. Occasionally the tablet computers being used to conduct the interviews would not register the GPS location at a specific home but instead would use the GPS location at the next available point of cellular or Wi-Fi connection.

Second, half of the wards were very rural, spread out, or hilly based on the topography of the Kathmandu Valley[14]. Thus, the sampler may have had to walk longer distances to identify eligible households. The measurement of distance is as the crow flies instead of based on roads that often wound around hills or other steep or challenging terrain. In these more rural wards, twenty (n=20, 13 percent) of the households reported a GPS location in a ward that did not match the expected ward of the originally sampled point. Often, these households were just on the other side of a ward boundary. These households might be a cause for concern as bigger distances between the originally-sampled buildings and the household may include more risk to validity. However, the representation of the exact ward boundary maps used in this study may have been slightly different from the socio-political reality on the ground. For example, one field researcher reported during debriefing that the head of the household claimed to reside in a ward that was different from the ward we had identified in satellite imaging due to recent local political changes in the area. It is unclear if our categorization or the head of household was more accurate in this particular situation. Regardless, this technicality does not factor into the analysis of any survey statistics.

Finally, the remaining eight (n=8, 5 percent) households were interviewed at a GPS location in the expected ward of the originally sampled point, but the distance between the survey GPS was closer to a different originally sampled point than the one the sampler identified. That is, the household may have been identified from the sampler's walk from one originally identified point, but the household was actually located closer to another point that the sampler would also visit. In some areas of Kathmandu, the houses were very dense and resulted in many buildings on OpenStreetMap identified in a very small area. This increased the probability that these buildings would be randomly selected in these areas. Therefore, it is possible that some of the households would have been associated with multiple sampled points. Because not every sampled building yielded eligible households with children 12-17 years old, the path the samplers walked to identify eligible households and the decisions they made resulted in overlap between sampled points within a ward that were close together. This is a threat to the validity of the sampling list because such households would have a higher probability of being included in the sample.

[14] This was true for two wards located in one municipality and three wards located in the other.

Figure 2.5 shows an example of the sampling process in Nepal using the semi-random walk. This figure is a stylized representation of the sample process and does not represent actual GPS coordinates of the sampled households or buildings in the ACV study. The red triangle represents a sampled building. The sampler would go to the location of these red triangles and start the pre-screening process. The blue circles represent locations where the surveys were conducted. In the ACV pilot, the majority of surveys occurred in the household of the eligible respondents. Therefore, the blue circles usually represent the location of the households identified in the sampling process. Cluster "A" shows a sampled building and six interviewed households that are close to the building. Cluster "A" represents a situation where the sampler followed the protocol and identified eligible households in close proximity to the particular building, as found in 76 percent of the sampled households. Cluster "B" and "C" demonstrate that a household selected from the second-stage sample process from original point "B" may have indeed been close to original building "C", as found in 5 percent of the ACV sampled households. The household with an (*) represents an interview that challenges the assumption that households were identified from only one originally sampled building. Cluster "D" represents a sampled point that resulted in households that were spread out due to lack of eligible households, rural landscape, or difficult-to-identify buildings, as found in 13 percent of the sampled households. Cluster "E" represents a sampled point that may have crossed a local boundary, as found in 5 percent of ACV sampled households. The household with (**) falls to the outside of a major road, represented by the yellow line. If this road marked a boundary change, it is possible household (**) is located in a ward not sampled, despite being identified based from the sampled building "E". Finally, the household (***) at the bottom of the map shows a household that was interviewed far from any originally sampled point. This represents households that were interviewed in cafes, community centers, or places of work upon request. This information was documented in the debriefing notes that I wrote at the end of each workday.

2.5. Metadata and Paradata Collected during the ACV pilots

The description of the sampling processes (in the above three sections) of the three ACV pilots provides transparent and realistic details not otherwise found in most published academic work. The social and logistical challenges faced in Tanzania and Nepal created situations that likely impacted the coverage and sampling error [Figure 2.1] in the ACV pilots, though, as mentioned before, it is not possible to quantify the extent. But establishing a sampling framework

and creating a sample are only the first steps of research. Up until this point, I have described the process of establishing a sampling frame and conducting a sample within a target population. The actual work of identifying, contacting, and surveying sampled respondents takes up the majority of field work and is arguably even more prone to social and logistic challenges and potential for error. Data collected during field work can be used to quantify the third type of representation error established by the Total Survey Error framework: nonresponse error. Before I discuss nonresponse error, I provide detail on the types of data collected during the sampling and data collection processes. This data is the foundation for analyzing the nonresponse error on the Total Survey Framework.

The ACV pilots contracted with local survey organizations, but oversight and on-the-ground management was done by me and the project PI. Through daily debriefing and constant personal communication, we were directly responsible for the training, management, and coordination of the field team. We made decisions about challenges and issues that arose daily from the beginning of the sampling process through the end of the data collection[15]. In order to have proper records of every step of the process, we collected paradata about the sampling and surveying process. As a reminder, *metadata* is defined as data about the data. It includes the survey tools, sample design, training materials and research protocols developed in advance to be used by the field team to collect data. *Paradata*, as defined by Couper (2005), is data about the process of data collection. The paradata from the ACV pilots is documented in three main sources: sample lists produced by the sampling team, data collection tracking of complete and incomplete interviews, and daily debriefing notes.

The sample lists created by the sample team in rural Tanzania consist of handwritten lists of names produced with the assistance of the sub-village leaders. Other summary information about the population sizes of the sub-villages was documented during the conversations with the sub-village leaders and by manually counting the names written on the sub-village lists. Due to privacy and ethical reasons, we were not able to photocopy, take photos of, or replicate the sub-village lists for our records. As noted above, it was not necessary to refer back to the original lists except in the cases where some of the summary information was missing. Key pieces of information included the total number of households, number of eligible households, and number of sampled households.

[15] In rural Tanzania, the field work period was 13 days for both the sampling and data collection. The urban Tanzania pilot scheduled 10 days for the sampling and 20 days for data collection (including rest days). In the Nepal pilot, sampling took approximately 3 weeks and data collection lasted for 20 days.

In urban Tanzania, the sampling process was formalized to include worksheets to be filled out by the *mitaa* and cell leaders. On these worksheets, the total number of cells (along with an estimate of the number of households in each cell) and the number of eligible households in a cell provided information used in constructing sample weights and population estimates, described in later sections. The worksheets provided the sampling team with a consistent protocol to use in the *mitaa* and helped the sampling team verify that all of the summary data was collected properly. These worksheets were entered by hand into Excel. In Tanzania, we also obtained published population statistics from local government offices.

Similar worksheets were used in Nepal to document the sampling process. Sampling in Nepal required intensive documentation of the number of buildings visited, the number of eligible and ineligible households in each building and estimates of household as we were relying on the sampling process for all information about population size in the wards. This information was compiled by the local team into a spreadsheet.

Tracking the data collection followed a similar process in the three pilot locations. At the end of every day in the field, the field team gathered together to debrief. During these sessions, the team reported on the households visited. Households were marked as "complete" if both the adult interview and child or children interviews were completed. We recorded whether the team needed to return to the household to conduct one or more interviews. In subsequent debriefings, we amended the documentation to show if the household has been completed or not. Households that were contacted or unavailable were indicated as attempted contacts. Households that were not completed sometimes included details about why the interview was not conducted – child away at school or nobody answering the door—or if the household refused or otherwise indicated they did not want to participate. This spreadsheet for tracking the data collection was updated daily to include details about the overall progress of identifying, contacting, and interviewing households.

Finally, qualitive debriefing notes provided additional paradata about the data collection process. These notes included more details to compliment the data collection spreadsheet. Data collected from the survey process was validated based on the information from the debriefing notes. This included situations where interviews were eliminated from the data or data was edited based on new information. For example, the age of a child reported in the household survey was occasionally incorrect and modified in the data to reflect the age given by the child during the interview with the child. Another situation involved a girl who was interviewed twice by two different members of the field team, though nobody was entirely sure why she did not mention to

the second interviewer that she had already participated. Her second interview was removed from the data. These are examples of information documented in the debriefing notes that was used to verify and edit the quantitative data and paradata.

The documentation of paradata is complex and nuanced. In a large-scale survey organization, paradata is standardized and collected automatically within most survey data collection software. For example, the SurveyToGo software used by the ACV project has the capabilities to track time spent on each question, patterns in responses, or retrospective changes to the data by the interviewer; any of which might suggest falsification or manipulation of responses in ways that threaten the validity of the data. In a small-scale survey such as ACV, the entire team could gather daily to discuss challenges and questions. Being on location with the field team meant that the PI and I were on hand to answer questions and problem-solve immediately. The small nature of the pilots and the hands-on supervision by myself and the PI generated a more organic and homegrown method of tracking paradata. That being said, a systematic way of documenting paradata would eliminate many of the challenges I faced when reconciling, cleaning, and properly classifying the paradata presented in this analysis.

In the ACV pilot, three categories of paradata were used collectively to construct measurements often found in publications: response rates and other outcome metrics, sample weights, and population estimates. All of these measurements require accurate data. Each of these measures will be described in a section below; section 6 discusses response rates, section 7 details the calculation of sample weights, and section 8 estimates population counts. I will describe how each of these measurements identifies and quantifies nonresponse error in the data collection process, the third type of representation error in the Total Survey Framework alongside coverage error and sampling error. I will not go into detail about sources of potential error resulting from the interview process itself such as interviewer effects, processing error, bystander effects, social desirability and acquiescence, and cultural power dynamics between respondents and interviewers, as there is a large literature on these topics in survey methods and psychology. Instead I will focus on how the sampling process and data collection can yield measurable outcomes often reported in journals.

2.6. Response Rates and Outcomes Measures using ACV Paradata

In published research, the relationship between nonresponse error and the data collection process is most commonly reported as *response rates* as a way for researchers to provide evidence of validity of the survey results. This is such common practice that journals and reviewers adapted formal and informal interpretations of how high a response rates should be in order to signify a "good quality" study (Carley-Baxter et al. 2009; Johnson and Owens 2002; Bennett et al. 2011). Yet, few journals establish clear guidelines on what exactly defines a response rate and what threshold, if any, meets the standards of "high enough".

Broadly, *response rates* represent the proportion of respondents or participants out of the sampled or target population (Kviz 1977). For example, the number of paper surveys returned divided by the total number of surveys mailed or the number of respondents who answered their phone or door (and completed the survey) when contacted. Historically, reporting response rates when discussing survey research is a standard that was elevated by public opinion polling (Marton and Stephens 2001). Readers of public opinion polls demanded transparency in the validity and representativeness of the survey and the sampling process. An opinion poll that asked a few purposefully selected individuals could not be trusted compared to a poll that surveyed a large number of people across the entire population. Ideally, all individuals who are sampled for a survey would respond to the survey, thereby resulting in a 100 percent response rate. Achieving a 100 percent respondent participation preserves the assumptions of a probabilistic sample being draw randomly from the target population. The more people who responded, the less nonresponse error in survey statistics resulting from differences between the type of people who responded compared to the type of people who did not respond. This means that your sample is more likely to be representative of your target population.

Complete (100 percent) respondent participation is rarely accomplished in practice. But high response rates are seen by journals, reviewers, and readers as a proxy to indicate low nonresponse error[16]; though evidence suggests this may be a flawed assumption (Groves and Peytcheva 2008). *Nonresponse error* occurs when there is a nonrandom difference between respondents and non-respondents that were on the sampling frame [Figure 2.1], resulting in potential nonresponse bias in survey statistics (Galea and Tracy 2007). The pressure placed on

[16] One example of an explicit expectation of response rates by the *Journal of Pharmaceutical Education* states that studies should have a 60 percent response rate, or 80 percent if the study is among college students (Fincham 2008).

authors and researchers to publish high response rates established the standard that only high response rates were acceptable[17]. In a paper by Carley-Baxter et al., the authors argue that there is not a clear understanding among academics of what it means to have "acceptably high standards":

> Anecdotally, some of our colleagues hold fast to the perception that it is harder to get studies published if they fail to achieve acceptable response rate standards. However, these same individuals readily admit that they do not have an accurate picture of what, if any, standards regarding data quality or survey error are imposed by journal editors when considering manuscripts which report results based on data analysis of surveys. (2009)

With an incentive to publish, author may seek to present their study favorably and find ways to report outcomes of the survey that depict a higher response rate and omits any concerns about data quality. For example, the inclusion of partially complete surveys or the exclusion of certain parts of the sampling frame (respondents for whom there was no additional information confirming eligibility) may inflate response rates.

In an attempt to standardize reporting response rates, the American Association of Public Opinion Research (AAPOR) published a report that provides a clear standard for survey methodologists to use for calculating and reporting four outcome measures: response rates, success rates, refusal rates, and contact rates (American Association for Public Opinion Research 2016). Each of these different outcome measures will be discussed and defined in detail in the following section. Such guidelines document what metadata and paradata metrics should be gathered during data collection regarding interviewed and non-interviewed units across different modalities of surveys, including phone interviews, web-based surveys, and household surveys. Using a standardized metric for reporting outcome measure allows comparisons between different surveys regardless of the sampling frames.

While not the first attempt to establish definitions (Smith 1999 as found in Carley-Baxter et al. 2009), AAPOR is the premier academic association for survey researchers working in the United States and a respected authority among survey research scholars[18]. The standards are clear, flexible, and adaptable to survey research in any field. Additionally, citing the AAPOR standards

[17] It has been well established that response rates for survey research are rapidly falling around the world leading to a number of dramatically titled research articles such as "Where Have All of the Respondents Gone? Perhaps We Ate Them" and "The End of the (Research) World As We Know It?"(Leeper 2019; Stedman et al. 2019). Response rates for mail in and phone surveys, the traditional method of data collection prior to the internet era, have fallen significantly since about the 1980s. Practitioners and academics race to find new methods, and the field of survey research is booming with new and innovative ideas for data collection, sampling, and measuring survey quality.

[18] While AAPOR is an American organization, the standards have been used in surveys conducted in other countries (Beerten et al. 2014)

allows researchers to clearly communicate the validity and quality of the data collected to reader, reviewers, and journals. Several journals now require the AAPOR definitions to be explicitly stated in published work, including the *American Journal of Pharmaceutical Education* (Reierson Draugalis, Coons, and Plaza 2008) and the *Journal of the American Medical Association* (JAMA). While the AAPOR standards for response rates and other outcome measures are well accepted by survey researchers, they are rarely used by academics in other fields.

2.6.1 Adaptation of APPOR Guidelines for 3 pilots

I have adapted the AAPOR *Standard Definitions: Final Disposition of Case Codes and Outcome Rates for Surveys Revision 9* from 2016 for the three ACV pilot studies. The AAPOR *Standard Definitions* are written as a tool that researchers can apply to any type of survey, regardless of the unit of observation, or sampling strategy. I specifically refer to the guidelines for "In Person Household Survey" (Page 23-27). I use the "household" as the unit of observation for consistency between the AAPOR definitions and the ACV pilots.

In this section, I summarize the standard definitions for four outcomes measures: response rates, cooperation rates, refusal rates, and contact rates. All of these outcome measures rely on a common understanding of household eligibility, contact, and completion of survey regardless of the sampling strategy used. I compare outcome measures across the three ACV pilots. Standardizing the response rate, and other outcome measures, allows for a standardized comparison potential nonresponse error in the ACV pilots.

AAPOR Definition of Eligibility

The foundation of all AAPOR outcome measures can be broken into the following categories: eligible households that were interviewed, eligible households that were not interviewed, households that were not interviewed and it is unknown if they would have been eligible, and ineligible households. These four categories are further divided into sub-categories, also described in Figure 2.6:

- Eligible Households that were interviewed
    - Completed Interview (I)
    - Partial Interview (P)
- Eligible households that were not interviewed

    - Refusals and break-off (R)
    - Non-contact (NC)
    - Not interviewed for other reasons (O)
- Households not interviewed and it is unknown if they would be eligible (UH)
- Ineligible households (IE)

Eligibility is determined based on the definition of the target population; the ACV pilots defined eligible households as households that included at least one 12-17-year-old who would be available to be interviewed. After the target population is defined and the sampling frame is established, the sample is drawn through any of the probabilistic or non-probabilistic methods available. Each sampled household must include one of the above categorizations at the end of the data collection process. The ACV debriefing notes document each interaction with the household and record the status of the household (eligible or ineligible) and the result of the final interaction with the household (interview completed or the reason the interview was not completed). If the household fell into the "unknown eligibility" category (UH), it may or may not have been possible to contact the household or no paradata was recorded about contacting the household, as depicted by dashed lines in Figure 2.6. I will apply this categorization scheme to the ACV pilots, and present examples, in greater detail below.

AAPOR Definition of Response Rates

APPOR outlines four different types of rates that should be included in all survey research based on the categorization of sampled households described above: response rates, cooperation rates, refusal rates, and contact rates. Each of these four measures can be calculated multiple ways, as shown in Table 2.2.

Response rates are the most familiar outcome metric to most researchers. Generally speaking, a response rate is defined as the number of interviews divided by the number of eligible households from the sample list. The AAPOR report provides six definitions of response rates; three include only completed interviews in the numerator and the other three are comparable but also include partial interviews in the numerator. As partial interviews were not relevant to the ACV project, I only consider three response rates: RR1, RR3, and RR5[19]. In Table 2.2, all three response rates have identical numerators: completed interviews (I). The variation across the three

[19] RR2, RR4, and RR6 are the official titles for the response rates that include partial interviews in the numerator. I omit these definitions as the ACV pilots did not have any cases with partial interviews. In order to properly follow the APPOR definitions, I maintain the discontinuous numbering system for clarity.

response rates depends on how households where there was unknown eligibility are included or excluded in the denominator. In RR1, *all* households that were unknown eligibility (UH) are assumed to be eligible but uncontacted. RR1 is the most conservative estimation of the proportion of households responding. Response rates calculated under the definition of RR1 will be lower than other definitions. RR3 adjusts the denominator to estimate that not all of the unknown households (UH) would have actually been eligible if they had been contacted. The proportion, represented by the *e* in the formula for RR3 on Table 2.2, is determined by the researchers (AAPOR, page 62). This estimation for *e* must be explicitly described if RR3 is to be reported by researchers. I describe below how I estimate *e* in the context of the ACV pilots. The final response rate, RR5, drops the unknown eligibility households (UH) from the denominator, assuming that *none* of the unknown households would have been eligible for the survey. RR5 is the least conservative response rate calculated and produces higher response rates than RR1 or RR3.

AAPOR guidelines offer more nuanced understandings of how respondents participate in a survey. The cooperation rate and the refusal rate assess which households actively refused to participate while the contact rate highlights which households were not able to participate due to a non-contact by the field team or an inability to participate in the survey during the field work time.

AAPOR defines *cooperation rates* as "the proportion of all cases [households] interviewed of all eligible units [households] ever contacted" (page 6). In the ACV pilots, we calculate the cooperation rates at the household level only, not for the individual child interviews conducted within the household[20]. The cooperation rates include all of the interviewed households in the numerator over all of the households that were contacted and were either interviewed or explicitly refused to participate. Cooperation rates are defined two ways (see Table 2.2)[21]. COOP1 includes that completed interviews (I) in the numerator divided by the sum of the completed interviews (I) plus the refusals (R), and other reasons for not completing the interview (O). The category of "Other reasons" include households that were contacted but were unable to participate due to reasons such as not having a proper translator, the participant not being in good health, or religious holidays preventing participation. Reasons specific to the ACV pilots are described below. The second measure of cooperation rate, COOP3, does not include households

[20] Individual children can refuse to assent. But the overall household consent was needed for a complete interview of adults and children.
[21] COOP2 and COOP4 include partial interviews in the numerator; these are irrelevant to the ACV project.

categorized as being not interviewed for other reasons (O). Thus, COOP3 will be greater than or equal to COOP1.

As a companion to the cooperation rates, *refusal rates* are defined as "the proportion of all cases in which a housing unit or respondent refuses to do an interview, or breaks-off an interview out of all potentially eligible cases" (page 7). That is, the numerator accounts for the number of households that refused to participate or ended an interview early and requested to be removed from the study[22]. As seen in Table 2.2, there are three versions of refusal rates that share the same denominator as the three denominators found in the response rates (RR1, RR3, RR5). The different denominators account for what proportion of the unknown eligible households to include in the estimate. The cooperation and refusal rates may be useful to report if a survey has a large proportion of refusals or was unable to conduct interviews for other systematic reasons such as language barriers. However, in the ACV project, there were so few refusals that the cooperation rates were very high, and the refusal rates were very low.

The final outcome measure defined by AAPOR is *contact rates.* Contact rates measure if a household was contacted by the field team. As shown in Figure 2.6, the flow chart of the AAPOR categorization of household eligibility, contacted households are categorized as resulting in a completed interview (I), a refusal (R), or another reason for an incomplete interview (O). The three variations of contact rates – CON1, CON2, CON3—are defined in Table 2.2 as the total number of contacted households over all eligible households in the samples. Variation between the three contact rate definitions again comes from differential inclusion of the unknown eligibility households (UH) in the denominator. Contact rates are best reported when there may be systematic concerns about households that were not contacted (NC). In the ACV pilots, the reasons households were not contacted vary greatly and will be described below.

Researchers should clearly state which definition of outcome metrics are being used in published papers. For example, using RR5 assumes unknown households would be ineligible and thus yields a higher response rate than RR3 or RR1. This is especially important if there is a large number of households with unknown eligibility; it is important for readers to be able to decern what assumptions were made about the unknown households when calculating and reporting the response rate. It is not always necessary to define and report all four outcomes measures but

[22] This request follows the IRB process of continuous informed consent. In the ACV pilot, there were no cases of partially completed interviews that were broken off.

doing so provides the readers greater transparency of data quality and potential for nonresponse error [Figure 2.1].

In the next section, I apply the AAPOR *Standard Definitions* to the ACV pilots. I demonstrate how to apply the methodology to the paradata gathered during the sampling and data collections processes in rural and urban Tanzania and peri-urban Nepal.

2.6.2 Eligibility in the ACV pilots

Using three different types of sampling strategies, the sampling process in each ACV pilot resulted in a list of households that we assumed to be eligible for the survey. We made this assumption based on household information from the local leaders (in Tanzania) or pre-screening process (in Nepal) prior to the start of data collection. In the ACV study, the target population was households with an adult at least age 18 with at least one child age 12-17 living in the household. More specifically, in order for a household to be eligible for the study, the adult had to be *present* to give consent for the children to participate in the study and at least one child in the age range had to be *present* and available to be interviewed. Households that did not have an adult present or where all of the children in the age range were not available for interview were considered ineligible.

A benefit of the intense sampling process conducted by the ACV sampling teams [described in section 2.2, 2.3, and 2.4] was that ineligible households were often screened out prior to the start of the ACV data collection. Ineligible household do not count as a part of the denominator for any of the AAPOR outcome metrics, so this pre-screening process does not affect any of the rates[23].

Eligibility Concerning Boarding School Students

The original intention of the ACV pilots was to include households with an adult respondent age 18+ and resident children age 12-17. We anticipated that we would find a few households where there would be children who were away for part or all of the field work period and therefore not be available to participate in the study. In the first pilot study in rural Tanzania,

[23] AAPOR specifies four additional criteria for eligibility: (1) the selection of individuals within the household, (2) proxy respondents, (3) substitutions, and (4) status days. None of these criteria were applied in the ACV pilots in ways that changed the results of the outcome metrics of response rates, cooperation rates, refusal rates, or contact rates in a way that affects potential nonresponse error.

we were unaware that we had selected a village located in a municipality where all secondary school students attended boarding schools. This led to a high number of older children in the village being away at school; we were unable to include many of these households in the study as there were often no other children in the age range at home. If the household included a boarding school student who was away but also at least one 12-to-17-year-old child who was at home, the household was eligible.

In the next study, we were more intentional in our screening process to only include households where children living in the household would be present during our field work period. In urban Tanzania, we also timed our field work period to be during school holidays, when we anticipated more students, even those who would normally have been away at boarding school, would be home. In Nepal, some children were visiting family, but fewer children attended boarding school away from their place of residence. Careful pre-screening in the urban Tanzanian and Nepal pilots decreased the number of sampled households later deemed ineligible because all eligible children in the household were away, compared to the rural Tanzanian pilot.

In order to assess whether boarding school students and other children away from their family during the field work periods potentially affected the nonresponse error, I constructed the AAPOR outcome measures two ways. First, households with children who were away were treated as contacted households who were eligible but not interviewed in the field work period. This falls into the specific category of non-contact (NC) because it was determined that there was someone eligible at the household, but as all eligible children were away, the interview could not be completed[24]. This is similar to if an adult were contacted but the interview was not able to take place because the adult was at work at all times that the field team attempted to visit the household; these interviews were not considered refusals because there was still an attempt to contact the household but without confirmation of recruitment as determined in the process of continuous informed consent. In the calculated outcome measures, this inclusion of the boarding school students as eligible, but not contacted is noted with the addition of BS (to stand for "boarding school" in Table 2.4) to the outcome measure label. In the second set of definitions, households with children away were treated as not being in the sample as they no longer fit the criteria of having children present in the household during the field work period; that is, households with boarding school children were considered ineligible (IE) and therefore not

[24] I consider this non-contact (NC) instead of not participating in the survey for "other reasons" (O) because the child being away meant that the child could not be contacted in order to start the assent process with the child.

considered in the construction of response rates, cooperation rates, refusal rates, and contact rates[25].

Other eligibility concerns presented in the APPOR Guidelines

In addition to the eligibility definition presented in Section 2.6.2, AAPOR specifies four additional criteria for eligibility criteria: (1) the selection of individuals within the household, (2) proxy respondents, (3) substitutions, and (4) status days. The *selection of individuals* within a household is a concern if a survey is designed to only interview one specific member of a household. *Proxy respondents* are defined as respondents who were not sampled, but who provide the information needed in the survey on behalf of the sampled individual. *Substitutions* can happen when additional households are added to the originally sampled list when a household is found to be ineligible or eligible but does not participate in the survey. And *status day* refers to a set time frame within which the data collection occurs.

Selection of individual: In many household survey designs, the target population includes only one adult per household (Smyth, Olson, and Stange 2019). In some surveys, like many national censuses, the household is represented by a self-appointed head of household. In other surveys, the adult is selected through a pre-determined protocol. Regardless of the protocol, nonrandom or random, individual members of the households are not known in advance. The ACV pilots sampled the household as a whole and did not attempt to identify a specific household member. The intended interviewee representative of the household was an adult women who was the mother or primary guardian of the children residing in the household (see below), but this role was occasionally filled by a father, grandmother, or other relative if the mother of the household was unable or unwilling to be interviewed.

Proxy respondents: Given the above definition of the selected individual, a proxy respondent is another individual who answers on behalf of the selected individual. In the ACV pilots, the predetermined protocol for interviewing household members specified that the mother or primary female guardian of the household's children should be interviewed, if possible. This was because adult females are likely to know more about the situation of children living in the household (Galdo, Dammert, and Abebaw 2019). In rural Tanzania, the mother was interviewed 69 percent of the time, and in both urban Tanzania and Nepal, the mother was interviewed 60

[25] For example, in Table 2.4 described below, RR1 for rural Tanzania is calculated with and without the boarding school children (BS).

percent of the time. In the remaining cases, a father, grandfather, grandmother, older sibling, or aunt was interviewed. This is not considered a proxy because the sampling unit was the household, not the individual adult in the household[26].

Substitutions: The AAPOR guidelines also recommend that it is important to report any substitutions, e.g., when a sampled household cannot be found and another is included in the survey instead. In urban Tanzania, one *mtaa* required substitution from the additional households on the sample frame when the originally selected households were deemed to be ineligible because the children were the wrong ages. In rural Tanzania, the quota system used within the sub-villages to determine proportional representation is not considered substitution. In Nepal, no substitution occurred; however, several households were added to the list when they were discovered in a previously selected building. The protocol when establishing the sample list was to include all eligible households in a building. Therefore, the discovery of additional eligible households in a building was not substitution but rather the addition of a household that *should have* already been on the list.

Status Day: The ACV study did not have a set *status day* that determined eligibility of households. The final categorization of completion, contact, or unknown el1igibility in all three pilots was determined by the last contact, per AAPOR recommendations (page 11).

2.6.3 Categorization of Contact and Completion in the ACV pilots

I used paradata collected during the sampling and data collection process to assign each household in the three ACV studies an AAPOR categorization of eligibility, contact, and completion as found in Figure 2.6. For some households, the appropriate category was obvious. For example, if a household completed both the adult interview and at least one eligible child interview, the household was considered to be a complete interview (I). For other categories such as non-contact (NC), I highlight several detailed examples that resulted in the NC categorization. In this section, I present examples of the four main categorizations of eligible households that

[26] Each household completed at least two surveys. One that was answered by an adult in the household, described above, and at least one answered by each eligible child. Each child (age 12-17) in the household was the respondent for the survey intended specifically for children age 12-17. If an adult had answered the child survey on behalf of a child, that would have been considered a proxy response. But this was not allowed in the ACV pilots; therefore, proxy reporting is irrelevant for the adult survey and not possible for the child survey.

were interviewed (I), eligible households that were not interviewed (R, NC, and O), households with unknown eligibility (UN), and several specific cases of ineligible households (IE).

The easiest to measure were households that had complete interviews (I). This meant that a household had a completed adult survey and a completed survey at least one eligible child interview. Households were considered to be complete once they achieved this status even if there were other eligible children who were not surveyed[27].

Households that were determined to be eligible but resulted in no survey being conducted are further categorized as refusals (R), non-contact (NC), or other (O). In the ACV pilots, a household could be considered eligible if the field team was able to confirm the ages of at least one of the children residing in the household fell in the age range. In most cases, this determination was the result of a short phone call to the household, a brief visit and conversation with some member of the household, or a conversation with a knowledgeable neighbor or community leader.

Following IRB protocol, the main adult participant consented to the study and gave consent for all eligible children in the household. Refusal (R) to participate could happen in one of two ways. First, an adult member of the household refused over the phone during the initial contact and the household was not visited at all. Second, an adult in the household refused only after the field workers were at the home. We did not have any situations where a respondent, adult or child, refused to continue once already starting with the survey process.

Defining non-contact (NC) in the ACV pilots was more nuanced. Sometimes, a phone call or visit was made, but the household was not available to participate in the survey. Reasons included being too busy, not being home on the specific day that the interviewers could visit, or the adults were at work and wouldn't be home within a reasonable time for the field research team to visit. It is possible that respondents may have invented excuses to avoid flatly refusing to participate due to cultural norms against direct refusals. There is a sizable literature in survey research about how different cultural norms about social expectations vary between countries (Johnson et al. 2002; Lalwani, Shavitt, and Johnson 2006). The nuance of these conversations with household members was not recorded during field work, so it is impossible to accurately

[27] The AAPOR definitions also specify "partial interviews" (P). In the ACV study, respondents were allowed to skip questions, but this situation is considered a complete interview despite the missing information.

categorizes these potential households as refusals (R) instead of non-contact (NC). Therefore, they are considered non-contact (NC) in the following analysis.

Other situations arose where a household was confirmed to be eligible, but the household did not participate in the interview (O for other). For example, in one case the sole eligible child in a household was experiencing mental health issues and could not knowingly assent to participate. Another situation involved a family that was in mourning and the field researcher determined that it would not be appropriate to ask them to participate. An interview that could not occur due to language barriers would also fall in this category, but this did not happen in the ACV pilots.

There were some cases, in all three pilots, where the field team completed the household survey with the adult but were unable to complete a survey with an eligible child in the household. These cases are included as being incomplete in "other reasons" (O) as this situation generally occurred when the child was too busy to be available during the field work period. We removed the completed adult survey from the data as a household needs both the adult survey and at least one survey of an eligible child in order to be considered a complete interview (I). This also applied in one situation where the child was interviewed with the consent of the adult, but the adult interview could not be scheduled within the field work period.

The next major categorization of households in the ACV pilot was when the field team was not able to determine whether the sampled household was eligible (UH for unknown household eligibility). The documentation in the paradata was not standardized across the three pilots, though similar language was used. In Table 2.3, I describe some of the common situations found in the paradata. These included "no attempted", "Household unsafe or unable to reach", "Unable to locate", and "Unable to make contact via phone". All of the households described in Table 2.3 are considered in the analysis and construction of AAPOR outcomes measures as unknown eligibility (UH).

Finally, the paradata documentation from the original sampled households also recorded specific households being eliminated from the study for being ineligible (IE). The most common reasons for excluding households from the sample was that the children in the household were the wrong age. We attempted to mitigate this situation through the pre-screening process but came across many households with 11-year-olds and 18-year-olds due to the local leader or other

members of the households not knowing the child's exact age during the pre-screening process[28]. If an adult in the household was interviewed and it was later discovered that the intended child respondent was ineligible, the household was excluded from the data.

2.6.4 Results of ACV Response Rates, Cooperation Rate, Refusal Rate, and Contact Rate

Once all of the ACV households are categorized, I applied the AAPOR methodology and definitions to construct the four outcome measures: response rates, cooperation rates, refusal rates, and contact rates. The standard categorizations allowed comparison of the outcome measures across the three pilots. For each of the three pilots, I calculate each of the four outcome measures according to all of the definitions presented in Table 2.2. This allows me to compare within-pilot variations resulting from different definitions of outcome measures.

As mentioned before, I also calculate all of the outcome measures including and excluding the boarding school students; first, excluding households where the child or children in the age range were absent from the household throughout the data collection period, thus categorizing the household as ineligible (IE) and then including households with boarding schools kids to be considered eligible but not contacted (NC) or interviewed – indicated by BS in the columns in Table 2.4. In the situation that considers boarding school children eligible, households with *only* boarding school children designates the boarding school children as non-contacts and therefore these households always included in the denominator of the outcome measures[29].

Table 2.4 reports the percentages for the three response rates (RR1, RR3, RR5), two cooperation rates (COOP1, COOP3), three refusal rates (REF1, REF2, REF3) and three contact rates (CON1, CON2, CON3). The methodology for calculating these rates is found in Table 2.2. The outcome measures are aggregated for the entire pilot due to small samples sizes for the finer geographic units: sub-villages in rural Tanzania, *mitaa* in urban Tanzania, and wards in Nepal.

In all of the formulas that require an estimate of how many unknown households would have been eligible (RR3, REF2, CON2), I calculate *e* as the inverse probability of being ineligible among the households in the original sample frame. In rural Tanzania, 25 percent of the

[28] When the field team made the initial contact with the households, we had a predetermined birthdate period that determined eligibility. In one case in Nepal, we made a single exception where a household had multiple children and one of them was turning 12 on the day we conducted the interview, and we felt it was unfair to exclude her from participating when she was officially the age we had asked for and her siblings were participating.

[29] If a household had at least one child at home who was able to participate in the survey, even if other children were at boarding school, the survey was considered complete (I).

households on the original sample list of households established with sub-village leaders were deemed to be ineligible upon contact. Using this information, I assume that 75 percent of the households with unknown eligibility due to non-contact would have been eligible (e=0.75). In urban Tanzania I estimate that 98 percent (e=0.98) of households would have been eligible and in Nepal 80 percent (e=0.8) of households would have been eligible. The higher estimates of eligibility in urban Tanzanian and Nepal are partially due to improved sampling strategies and pre-screening process employed in these pilots.

Across all pilots, response rate 1 (RR1) is lower than response rate 3 (RR3) and response rate 5 (RR5). RR1 assumes all households with unknown eligibility (UH) would be eligible and thus are included in the denominator [Table 6.1.2]. Across all three pilots, response rates (regardless of the exact definition used) were between 64 percent (RR1 in rural Tanzania when including households with boarding school students) and 91 percent (RR5 in rural Tanzania excluding households with boarding school students). The inclusion of households with boarding school students in the denominator reduces the response rates across all definitions and across all pilots.

Rural Tanzania has the largest range of possible response rates largely driven by the status of households with boarding school students and a large proportion of households that were unknown eligibility (UH). The ranges of response rates in urban Tanzania and Nepal are narrower than in rural Tanzania; urban Tanzania response rates ranged between 73 percent (RR1 with boarding school students) and 84 percent (RR5 without boarding school students) and in Nepal the range was 76 percent to 86 percent for the same minimum and maximum definitions.

The APPOR *Standard Definitions* report (2016) does not make specific recommendations about which response rates to report in published works so long as the authors are explicit about which response rate is reported. The purpose of reporting a response rate is to communicate the potential for nonresponse error in the survey. Overall, the response rates in the ACV pilots are high and fairly constant across all definitions (RR1, RR3, RR5). If required to select only one definition to report, I would recommend using the RR3 definitions as the results fall in the middle of the extremes of RR1 and RR5. RR3 utilizes the paradata to estimate of the proportion of unknown households that *would have been* eligible. This data driven approach best captures the nuances the sampling and data collection process and the social and logistical challenges faced by the field team to identify households.

Deciding whether to report the response rates that exclude or include households with boarding school students (BS) would depend on if there could be nonresponse error correlated specifically to households that sent children to boarding school and households that did not. For example, an analysis of family wealth or educational attainment may be sensitive to nonresponse error of the households that were not included in the survey because all eligible children were away at school. But for most analysis of the ACV pilots, the high and generally consistent response rates suggest that nonresponse error may be minimal.

To support the results presented in the response rates, I also report the cooperation rates (COOP1 and COOP3) and refusal rates (REF1, REF2, and REF3) in Table 2.4. The *cooperation rates* report the number of interviews over the number of households contacted and the *refusal rates* report the number of refusals over the number of households [Table 2.2]. Across the ACV pilots, the cooperation rates were very high; households that were contacted were very likely to participate in the study.

Rural Tanzania had near universal cooperation and no refusals. The refusal rates in urban Tanzania were also very low; regardless of the definition used (REF1, REF2, or REF3) the refusal rates in urban Tanzania were 4 percent. It is possible that the high cooperation rates and low refusal rates in Tanzania resulted from the team being accompanied by a local leader who conducted the introduction between the field team and the household. The partnership with a local leader may have increased the legitimacy of the field team, so households were more willing to participate in the survey (Groves, Cialdini, and Couper 1992). Alternatively, households may have felt more social pressure to participate because of the presence of the local leader. In the debriefing notes recorded daily, such social pressure was not reported by the field team; the field team followed the informed consent protocol that assured adult household members that participation was voluntary.

In Nepal, the cooperation rates and refusal rates were not at all affected by the inclusion or exclusion of households with boarding school students. The cooperation rates (particularly COOP1) was lower than in the Tanzanian pilots—88 percent compared to 94 percent (urban) and 99 percent (rural). In Nepal, the sampling team and data collection team conducted work without the involvement of local government officials. The entire team had name tags clearly identifying them as being part of a local organization, but the higher refusal rates and lower cooperation rates reflect the challenges and extra effort the team had to make to explain the project and engage participants in informed consent interactions compared to the Tanzania pilots.

In the ACV pilots, the cooperation rates and refusal rates calculated are functional inverses of each other. They each convey a similar message that household participants overwhelmingly cooperated in the ACV study if the field team was able to contact the household. High cooperation and low refusal rates suggest that nonresponse error caused by a potential difference between households that refused to participate and households that did participate is small. As with the response rates, AAPOR recommends that researchers communicate to readers which definition of the cooperation rate and refusal rate was used if such a rate is reported in a published article. For the ACV pilot, the preferred measure reported would be REF2 as it again uses the paradata driven approach to appropriately account for households with unknown eligibility. The preferred cooperation rate would be COOP1 as it accounts for households that did not complete an interview for reasons other than refusals (O); this was a very rare occurrence in the ACV pilots as described in section 6.3.

The final outcome measure recommended by AAPOR is the contact rate. The *contact rates* convey the success of the field team in contacting and determining eligibility of sampled households, as represented in the flow chart in Figure 2.6 (I + R+O). In Table 2.4, the contact rates (CON1, CON2, CON3) for all three pilots report similar percentages to the response rates. The contact rates indicate that between 77 percent and 96 percent of eligible households were contacted. In the ACV pilots, the contact rates reflect the hard work of the field team to find sampled households from the sample lists. In Tanzania, the help of local leaders was an essential element in identifying eligible households and making introductions. In Nepal, the contact rates were slightly higher – 90 percent (CON1 without boarding school students) to 96 percent (CON3 without boarding school students) – compared to Tanzania pilots. The Tanzanian rural pilot had contact rates of 77 percent (CON1) to 90 percent (CON3); and they were 82 percent (CON1) to 89 percent (CON3) in the city. The extensive sampling and pre-screening process in Nepal functioned as the first contact with household; therefore, when the field team called or visited the household during data collection, the sampled household had already talked to a member of the field team. In many households, the adult respondents remembered the interaction with the sampling team and had been waiting for a member of the field team to call to set up a time to conduct the survey. These high contact rates in Nepal suggest that pre-contact with households during the sampling process could increase contact rates during data collection[30].

[30] The sampling team was trained to use materials that reflected the IRB process of recruitment and informed consent. The informed consent process started with the first contact with the household and was continued during the data collection process.

The inclusion of households with boarding school students (BS) influenced the contact rates differently than the other outcome measures. For each of the pilots, CON1 and CON2 were higher when boarding school students were included (columns with BS) compared to columns that excluded the households with boarding school students. This is different from all three definitions of response rates; the inclusion of boarding school students decreased the rates for RR1, RR3, and RR5. The mathematical mechanisms at work in the formula balance the number of households with unknown eligibility and the non-contact households (NC) in the denominator. In Table 2.2., the formula for CON3 does not include these unknown households as it considers that none of the unknown households would have been eligible if they had been contacted (similar to RR5 and REF3). The inclusion of boarding school students decreases the CON3 rates in all three pilots.

Why does this even matter? The inclusion or exclusion of households with boarding school students, even in small samples like the ACV pilots, can change the outcome measures to look more or less favorable. This variability could be utilized to manipulate the results of a small study to report highly favorable results in order to increase the possibility of getting published. This manipulation, while technically accurate, masks the complexity of the fieldwork and decisions of eligibility criteria, recruitment, and sampling. When the study is small, which outcome measures are reported may not affect the survey results or increase potential error (coverage, sampling, or nonresponse error as established by the TSE framework). In a large study or a study where there is a large sub-population of participants that could be defined in multiple ways, such decisions could be highly influential in the results. Therefore, it is important for authors and researchers to clearly state how these sub-populations are or are not being included in the calculation of outcome measures, including response rates and contact rates. The importance of considering the sub-population is necessary if there is a potential for nonresponse error affecting specific variables of interest.

The primary purpose of reporting outcome measures is to quantify the sampling data collection process in a way that indicates that the results of the survey are not biased by nonresponse error. It is unlikely that all four of these types of rates would be included in a published paper as they are all intricately related. This application of the APPOR *Standard Definitions* to the ACV project demonstrates how different outcome measures and different definitions of outcome measures can be constructed through varying interpretations of the end result of household visits and interviews in paradata and documentation of the data collection process. Reporting outcome measures such as response rates without properly defining the

method of calculating the rate is misleading. Researchers should be transparent and specific when reporting outcome measures and highlight any potential nonresponse error due to refusals, noncontact, or unknown eligibility households.

Journals and reviewers should also reconsider rejecting manuscripts based solely on small response rates or high refusal rates. Small sample sizes in the ACV pilots result in variability of each outcome measures. A single household refusal may increase the refusal rates substantially in a small-scale survey without contributing nonresponse error to the overall study. Nonresponse error is an issue only if there is a correlation between households that do not respond and the variables of interest. Low response rates do not *prove* there is nonresponse error just as high response rates do not suggest a perfectly bias-free survey. The best practice for any survey is to report clearly which outcome measures were calculated and provide other analyses of nonresponse error, such as constructing weights and population estimates, described in the next two sections.

2.7 Constructing Weights from ACV Paradata

Weighting in an important part of any probabilistic sample as it allows users to adjust the results of outcomes measured by the survey to represent the target population by accommodating sample design and nonresponse (Kalton and Flores-Cervantes 2003; Yansaneh 2003; Solon, Haider, and Wooldridge 2015). Sample statisticians calculate weights in order to adjust the results of sample statistics so that they more accurately reflect population parameters. The construction of weights for the ACV pilots allows for a detailed look at the usefulness of weighting data.

Samples weights have two main functions. First, weights adjust the sample to reflect the descriptive size and composition of the target population. This can be useful for researchers working with raw frequencies. Second, weights can be applied to analysis to adjust the specific sample statistics to reflect the size of the underlying target population (Gelman 2007). It is only necessary to apply sample weights to an analysis if there is a concern that coverage, sampling, and nonresponse error may impact the results of the particular variable of interest (Makela, Si, and Gelman 2014). In this section, I describe the process of creating weights for each of the three pilots in the ACV study. I demonstrate the usefulness (or not) of applying sample weights to adjust the sample to represent the descriptive target population size. Finally, I conduct an analysis showing the application of sample weights for a variable of interest in the ACV pilots that is

potentially related to nonresponse error. I conclude this section by identifying potential reasons that researchers should or should not utilize survey weights.

Typically, there are three different elements to sample weights. A *base weight* (also called a sample weight or a design weight) takes into account the sampling process and allows the sample to be scaled to the size of the target population. For example, a random sample may include approximately 10 percent of the total households in a population and the base weights can be applied so that each household represents 10 other (unsampled) households when researchers present descriptive statistics. A *non-response weight* adjusts for the nonresponse bias possible in the sample. For example, households within a geographic area that have responses can represent households in that same area that didn't respond if the researcher assumes that all households in the area share homogenous characteristics, and therefore would have answered similarly. Finally, a *post-stratification weight* allows researchers to re-calibrate the sample to look more similar to the established target population. For example, if immigrant households are underrepresented in a sample, these weights would "scale up" or overrepresent the sampled immigrant households to accurately represent their proportion of the target population. In this exercise, I have created base weights and nonresponse rates. The product of the base weight and the nonresponse rate equal the *final weight*. I do not create post-stratification weights because we do not have accurate enough distributions of the target population of households with children 12 to 17 to make any adjustments. Additionally, we did not make any sample design choices that resulted in purposeful over- or under- sampling of a particular group[31].

The base weight (W_B) is calculated as the inverse probability of being selected (p):

$$W_B = \frac{1}{p}$$

I calculate a different base weight for each primary sampling unit (PSU) in each of the three samples: sub-village, *mtaa*, and ward. The process of calculating the probability of an eligible household being selected depends on the sampling process that happened within each pilot.

To calculate the nonresponse weight, I use the response rates calculated in the previous section. In order to explore the construction of weights fully, I calculate a maximum of six

[31] The missing boarding school student households is not something that can be fixed with post-stratification weights as we do not have population figures that show the number of children in boarding schools within our geographic areas. Thus, the boarding school students will be adjusted for in the non-response weights following the logic presented in the construction of non-response rates.

different nonresponse rates for each PSU in each sample based on the three different response rates (RR1, RR3, and RR5) and the inclusion or exclusion of the boarding school children for a total of six nonresponse rates. Overall, the nonresponse weight (W_{NR}) is the inverse probability of the specific response rate (*RR*)

$$W_{NR} = \frac{1}{RR}$$

Nonresponse weights for each of the different definitions of response rate are written as:

- W_{NR1} to refer to the nonresponse rate that uses the RR1 response rate
    - W_{NR1BS} to refer to the nonresponse rate that uses the RR1 response rate that includes boarding school students
- W_{NR3} to refer to the nonresponse rate the uses that RR3 response rate
    - W_{N31BS} to refer to the nonresponse rate that uses the RR3 response rate that includes boarding school students
- W_{NR5} to refer to the nonresponse rate that uses the RR5 response rate
    - W_{NR5BS} to refer to the nonresponse rate that uses the RR5 response rate that includes boarding school students

In all cases, the final weight (*W*) is the product of the base weight and the nonresponse weight:

$$W = W_B * W_{NR}$$

For each of the pilots, the combination of base weights (W_B) and six variations of the nonresponse weights (W_{NR}) defined above produces six different possible weights for each of the PSUs for each of the pilots. In order to make a recommendation of which particular weights to use, I present figures that show the distribution of the magnitude of the weights. Each weight represents the inflation factor to be used when conducting analysis about a particular parameter of interest. Thus, the weights can be interpreted as the number of eligible households in the PSU that are represented by each individual respondent household present in the sample.

2.7.1 Rural Tanzania

The sample in the rural Tanzanian village was almost a complete census of eligible households in some sub-villages. An accurate census with full participation would not require weights. As we almost achieved this, the weights for each household in each of the sub-villages

are not very large. For each sub-village, I construct the base weight from the number of households in the sampling list divided by the known or estimated number of households with 12-17-year-old children in the sub-village.

In three of the seven sub-villages, we did not have an accurate count of the target population households due to social and logistical challenges described in the sampling process. In order to construct weights, I estimated the total number of eligible households based on the proportion of households with 12-17-year-olds out of the total households in the remaining four districts where both figures were known; on average 29 percent of households in a sub-village included at least one 12 to 17 year old child. I established the denominator of target households, actual or estimated when missing, in order to calculate the probability of an eligible household being sampled within each sub-village. Thus, the base weight is the inverse of *p*, where

$$W_B = \frac{1}{p} \; where \; p = \frac{Sampled \; HH}{Total \; eligible \; HH \; in \; PSU}$$

These base weights (W_B) were combined with nonresponse weights (W_{NR}) calculated using the three different types of APPOR response rates described above ($W_{NR1,}$ $W_{NR3,}$ W_{NR5}). In addition, I calculated nonresponse weights at each of the three AAPOR rates ($W_{NR1BS,}$ $W_{NR3BS,}$ W_{NR5BS}) to account for the inclusion and exclusion of boarding school children for a total of six possible candidates for final weights (W). Figure 2.7 represents the spread of the six iterations of the final weights for each of the sub-villages. The orange dots represent the median final weight of the twelve variations calculated while the blue represents the minimum and the grey represents the maximum possible final weight. The range of possible final weights shows the extent to which the base weight and nonresponse weights vary because of differences in the original sampling frame and the response rates in each sub-village. Through these weights, we are able to generalize our sample findings to the target population of households with 12 to 17-year-old children in the selected village in Tanzania. We cannot generalize to other areas of Tanzania.

In the Tanzanian village pilot, the sample of eligible households was close to a census of the eligible households in each sub-village. Thus, the weights attached to each respondent household are barely larger than one [Figure 2.7]. In sub-village 1, each household has a weight of 1.18 (the overall lowest median) and the highest overall median (1.78) is in sub-village 7. Most of the sub-villages, particularly sub-villages 1, 2, and 5, have a narrow range. The sub-villages with large spreads (6 and 7) were also the sub-villages with higher than average proportions of eligible households in the sub-village (that is, more households with children ages 12-17) and higher than average numbers of households that were unknown eligibility due to non-contact. The

important lesson here is that when facing a small eligible population size, small deviations from the mean can result in great divergences in weights and non-response rates.

For this pilot, using the weights provides only marginal added value to the overall results from data collected from the sampling frame. But the overall adjustment to the sample in order to reflect the target population is fairly minimal and thus would not add great complexity or additional concern for error in the descriptive statistics. As we will see below, applying the weights increases the standard error in variables of interest compared to not using any weights at all.

2.7.2 Urban Tanzania

Weights constructed for urban Tanzania followed a process similar to the construction of weights for rural Tanzania as each stage of the multi-stage sampling process recorded the known probability of selection of the particular unit. The primary sampling unit in urban Tanzania is the *mtaa* and every *mtaa* had an equal probability of being selected. In total, 23 *mitaa* were sampled but two were later excluded for being too rural and one was used for training the field team.

According to our sampling process, cells are a secondary sampling unit. One cell was randomly selected from a complete list of cells for each *mtaa.* Remember, a cell theoretically represented ten households that were all known by a single local leader known as the cell leader. In practice, the cells in urban Arusha were larger and many had between 20 to 50 families. Finally, a complete list of eligible households in the cell was sampled from to create the sample list. The base weight is one over the product of the probability of the *mtaa* (P_M), cell (P_C), and household (P_H) being sampled in each *mtaa*.

$$W_B = \frac{1}{P_M * P_C * P_H}$$

Where

$$P_M = \frac{Randomly\ sampled\ Mtaa}{All\ Mitaa\ in\ Arusha\ Urban} = \frac{20}{122} = 0.164$$

$$P_C = \frac{Randomly\ sampled\ cell}{All\ cells\ in\ particular\ mtaa} \quad e.g. = \frac{1}{7}$$

$$P_H = \frac{Sampled\ HHs}{Total\ eligible\ HHs\ in\ Cell} \quad e.g. = \frac{10}{40}$$

In most cases, we knew the number of households only from the sampled cell. Thus, the calculation of weights assumes that other cells in the *mtaa* would have similar numbers of households in the other listed cells that were not sampled. This is a major assumption as it was not always clear if the lists created with the help of *mtaa* and cell leaders were complete. This potential issue will be addressed further in the discussion on population estimates below where I compare the assumptions of population size and distribution with the established population reports.

The base weights (W_B) were combined with the different six nonresponse weights (W_{NR}) calculated in exactly the same fashion as the rural Tanzania pilot to create the final weight (W). There are six total final weights candidates for urban Tanzania. Figure 2.8 shows the distribution of the minimum, median, and maximum values of the final weights for each of the *mitaa*. These weights allow us to generalize our sampling findings to all households with at least one 12-17-year-olds in Arusha. In comparison to the weights for the rural Tanzania pilot, the weights for urban Tanzania are larger; the average median final weight is 85.5 (min = 2.8; max= 234) and the spread of all of the final weights ranges from 2.6 to 341.6. A larger weight indicates that each sampled household represents a greater number of total households.

A significant factor in the size of the final weights is the small proportion of *mitaa* included in the sample out of the total *mitaa* in Arusha; this increased the size of the base weight for all *mitaa* which in turn increased the size of the final weight. The *mitaa* across Arusha vary greatly in population size. For example, *mtaa* 11 and *mtaa* 16 were both small neighborhoods of ethnic minority families that were very different from other areas of Arusha. In each of these two *mitaa*, there was only one cell, and the ACV sampling resulted in close to a census of the eligible households with 12-17-year-olds. Thus, respondent households in these *mitaa* have small final weights (median final weights at 6 and 2.6 respectively) because the *mitaa* had a small number of cells and an overall small population size compared to other *mitaa*.

In contrast, some *mitaa* have larger numbers of cells and large cell sizes (i.e. many households per cell). For example, *mtaa* 21 reported 14 cells and within the single cell that was randomly sampled, there were 29 eligible households. *Mtaa* 2 reported 11 cells, with 40 eligible households in the sampled cell. These two *mitaa* both have high median final weights (W = 164 and W = 122 respectively) as each household included in the study is weighted to represent large populations in these *mitaa*.

The calculation of the base weights remained stable for each of the *mitaa* as the estimates of base weights are produced in the sampling process, not the data collection process and identification of AAPOR categorizations. The differences in the range of possible final weights is driven by differences in the nonresponse weights and in the inclusion or exclusion of boarding school children in the calculations. *Mitaa* with consistent measures of response rates across the different definitions (RR1, RR3, and RR5) produce final weights that are identical across all twelve calculated final weights; for example, *mitaa* 4, 8, 9, 15, 16, 17, and 21 have no spread of final weights as shown by the minimum, median, and maximum values overlapping in Figure 2.8. The calculation of the base weight affects the size of the final weight – i.e. the final weight in *mtaa* 4 is 14 and the final weight in *mtaa* 8 is 116 —but the lack of spread indicates stability across the six nonresponse weights calculated. For these *mitaa*, the selection of final weights to use in analysis does not matter as they are all the same. In contrast, *mtaa* 23 has a wide range of final weights. This *mtaa* was highly influenced by several boarding school children and unknown eligible households (described in the above section as being households were the field team was unable to establish the eligibility status of the household). With these factors influencing the response rate calculation, the range of final weights is wide (minimum = 136.6; maximum=341.6). In mitaa with wide ranges, the selection of final weights used in the analysis depends greatly on the definition of response rate used.

Given the assumption made about the calculation of the base weights and the wide variation in nonresponse weights, it is difficult to know which final weights should be used in analysis of the urban Tanzanian households. Theoretically, the small sample (n=145 households) of the pilot could be weighted to reflect the population of over 100,000 households of Arusha[32]. But it is important to clearly define the calculation of weights and assumptions in order to assess if such weighting is valid and appropriate. The large magnitude of the weights and many assumptions required in the construction of the weights suggest it would be unwise to present descriptive statistics that adjust the sample size to the size of the target population. More on this will be discussed in the section on population estimates. Finally, just as in rural Tanzania, the use of weights should be carefully considered based on specific parameters of interest that might be affected by nonresponse error, as demonstrated below.

[32] Number of households (n=104,093) provided by Arusha District office.

2.7.3 Nepal

Due to the nature of sampling in Nepal, the process for calculating a base weight was vastly different than the previous pilots. The fundamental challenge in Nepal was that there was no way to accurately create a sample frame based on the target population. The sampling of households via satellite imaging of OpenStreetMaps inevitably included eligible and ineligible households. In the sampling process and response rate calculation, these ineligible households are ignored completely. Ineligible households would not be included in the sample at all and therefore are not considered respondents or potential respondents. However, when constructing weights, it is important to treat the eligible households as inherently different from ineligible households. We must consider the proportion of buildings identified in OpenStreetMap that *would have contained* at least one eligible household *if* they had been sampled. Of course, not all buildings would have an eligible household, given that the target population includes only households with 12-to-17-year old residents and not every building will include households meeting this requirement.

In order to demonstrate this difference, I have constructed base weights (W_B) three different ways. All base weights start with the same probability of a primary sampling unit (PSU) – in Nepal this is the ward – being included in the sample (P_W). These are stratified by municipality. We aim to be able to generalize to all households with a 12-17-year-old in two specific municipalities in the Kathmandu District of Nepal.

The next step of the sampling process involved the identification of buildings based on the satellite imaging. Each building has a known, non-zero probability of being selected. Remember that the sampling was conducted where all buildings in a ward were identified and randomly selected. The samplers visited the selected buildings and determined the eligibility of households. If there were not eligible households in the identified building (the starting point), the samplers continued to an adjacent building.

A preliminary base weight (base weight 1 or W_{B1}) establishes a weight of the number of buildings identified as a starting point divided by the total number of buildings in the ward. This maintains the probability sample of the starting buildings as being randomly selected. Of the eligible households (with 12-17-year-olds) in the selected building, they all have a 100 percent probability of being included in the sample as per the instructions to the sampler. The probability of a building being selected as a starting point over the total buildings possible to be selected as a starting point (P_{B1}) is the second part of the base weight 1 equation.

$$W_{B1} = \frac{1}{P_W * P_{B1}}$$

Where

$$P_W = \frac{Randomly\ sampled\ wards}{Total\ wards\ in\ municipality}$$

$$P_{B1} = \frac{Buildings\ selected\ as\ a\ starting\ point}{Total\ buildings\ in\ PSU}$$

However, not every building has a non-zero probability of being included in our sample because not every building was home to a household with a 12-17-year-old resident. The base weights constructed above highly inflate the weight of each respondent within a ward. On average, the sampler visited five buildings for every one randomly-selected starting building. This was done through a random walk. An additional complication is that the process of doing a random walk from a probabilistically selected starting point is no longer a probability sample (Bauer 2016).

In order to address the proportion of buildings that would be randomly selected as a starting point but would not yield an eligible household, I calculate two additional variations on the base weights. First, I account for all of the buildings visited by the sampler over the total number of buildings in the ward (base weight 2 or W_{B2}). This effectively inflates the numerator by five without changing the denominator[33].

$$W_{B2} = \frac{1}{P_W * P_{B2}}$$

Where

$$P_{B2} = \frac{Sampled\ building + All\ additional\ buildings\ visited\ by\ sampler}{Total\ buildings\ in\ PSU}$$

Alternatively, I estimate the proportion of buildings that *would have* an eligible household based on the known proportion of buildings with an eligible household visited by the

[33] The samplers visited approximately 5 buildings per originally sampled building in order to find the needed number of eligible households. This is reflected in this calculation of the base weight 2.

sampler (base weight 3 or W_{B3}). This strategy reduces the denominator and numerator to only include the probability of a randomly identified starting building with an eligible household.

$$W_{B3} = \frac{1}{P_W * P_{B3}}$$

Where

$$P_{B3} = \frac{Sampled\ HHs}{e(Total\ buildings\ in\ PSU)}$$

where *e* is an estimated proportion of eligible households per building in the PSU based on a calculation of the number of eligible households identified divided by the total households identified in the all visited buildings in the PSU. I then create an estimated average of eligible households per building across the entire ward based on the observed data.

Both of these methods (W_{B2} and W_{B3}) attempt to consider only the target population when creating base weights and as an added confirmation, the results from each of these corrections are very similar (and very dissimilar from the base weights constructed with no regard for eligibility). I calculate all iterations of the product of the base weights 2 and 3 with the nonresponse weights calculated using the same methods as in the other two samples, but only report base weight 3 here due to the high similarity with base weight 2.

The Nepal final weights (*W*) were calculated using the same process as in Tanzania to combine the base weight and the nonresponse weights (W_{NR}) calculated using the different AAPOR response rates ($W_{NR1,}$ $W_{NR3,}$ W_{NR5} and $W_{NR1BS,}$ $W_{NR3BS,}$ W_{NR5BS}). Similar to the Tanzanian pilots, this combination yields six different potential final weights that explore the three response rates with and without boarding school students. In Figure 2.9, I compare the six final weights produced using both base weight 1 [Figure 2.9A] and base weight 3 [Figure 2.9B]. The main difference between the two figures is the final weights produced using base weight 1 (W_{B1}) are significantly higher than the final weights produced using base weight 3. If we did not account for the proportion of eligible households per building in the sampling process, we would be applying weights that inaccurately describe the population of households with 12-to-17-year old residents. The utilization of paradata about the sampling process greatly improves the calculation of sample weights. Therefore, I recommend using final weights calculated by base weight 3 (W_{B3}) for the Nepal pilot analysis.

In Figure 2.9B, the final weights calculated for Nepal are similar in magnitude to the Tanzanian urban weights. This provides support for the validity of the weight calculations as the geographic area and starting population sizes of the two pilot sites are similar. The average median final weight across the wards is 67.8 (min = 30.9; max = 140). Like in urban Tanzania, the magnitude of the final weight is predominately driven by the base weight, in this case calculated by base weight 3. Similar to my recommendations for the urban Tanzania pilot, the large magnitude of the sample weights, instability of the results based on definitions, and assumptions made in calculating the weights, I would not recommend using weights to adjust the sample to reflect descriptive statistics of the target population in the Nepal pilot. This will be discussed in more detail in the section on population estimates.

The spread of the minimum, median, and maximum final weights is driven by the nonresponse weights (W_{NR}) in each ward. The Nepal final weights benefit from a lack of variation in the response rates because of the pre-sampling contact process where a team member confirmed eligibility in advance. This sampling process, as described above, led to high response rates and high contact rates. In two of the wards, there is no difference between the multiple response rates within the wards in Nepal resulting in fewer than 6 unique final weights. The wards with wider variation in final weights – i.e. T6, T7, and N3 – were wards that are located closer to Kathmandu City center. These wards had more varied nonresponse weights as these wards also had lower response rates due to refusals and noncontact of households as described in the above section. I will explore the relationships between response rates and specific variables of interest in the next section.

2.7.4 Application of Weights to Social Desirability Index

As demonstrated above, the calculation of final sample weights depends on the proper specification of base weights and nonresponse rates[34]. The base weight adjusts for the inflation of the sample size to the target population. Nonresponse weights adjust for the potential that survey respondents and non-respondent may have answered a specific question or set of questions differently. Transparency in this calculation is another way that researchers can provide evidence that their results are valid. The base weights produced in the two Tanzanian pilots used similar methodology of calculation. In Nepal, I recommend using the base weights 3 (W_{B3}) produced using the estimated eligible buildings as a proportion of total buildings. For all three pilots, I

[34] And post-stratification weights if this applies to a particular survey. It did not in the ACV pilots.

recommend using the nonresponse weight based on response rate 1 (W_{NR1}). Response rate 1 (RR1), as described in above, is considered the "minimum response rate" and is the most conservative estimation of response rate among households of unknown eligibility. Therefore, the nonresponse weights constructed using this response rate will provide a conservative weight. The product of the base weight and the nonresponse rate equals the final weight reported in the above section, though each of the weights could be applied to a sample independently.

However, weighting data is only meaningful if the desired outcome is to inflate the respondents to reflect the population figures or in the context of parameters of interest. In the first situation, the base weight will correct for the sample size to reflect the underlying target population. Given the ACV pilots' small sample sizes, it is unlikely that we would apply the base weights given the large magnitude of the weights, particularly in the urban Tanzanian and Nepal pilots. The magnitude of the base weight (and as a result, the final weight) leads to each household in the study representing an average of 85 households in the urban Tanzania pilot and 65 households in the Nepal pilot while each of these pilots only surveyed 145 and 155 households respectively. With all of the social and logistical challenges of the sampling processes, the coverage and sampling error threatens the validity of the sample weights representing such large populations. We conducted only one round of data collection in each of the populations. It is also not possible to test the validity our single sample in the overall sampling distribution.

Instead, we can turn to the second situation of applying the nonresponse weights to specific parameters of interest. The nonresponse weights are necessary if a question of interest may have been answered differently by respondents and non-respondents; this is often referred to as *unit non-response bias*. I test for nonresponse bias in the ACV data by applying different nonresponse weighting schemes and comparing the estimated means and standard deviations for unweighted and weighted estimates on a composite variable, described below. Specifically, I am testing the three nonresponse weights: $W_{NR1,}$ $W_{NR3,}$ W_{NR5}. This methodology, adapted from Blom (2009), purposefully selects variables that may be correlated to nonresponse. If the nonresponse weights are calculated accurately, then the application of weights to the survey can add value to the analysis of the specific parameter[35].

In the ACV pilots, we asked the adult respondent a series of questions to construct a social desirability index. The social desirability index used in the ACV pilot was constructed

[35] Blom (2009) applies this method to the European Social Survey which details nonresponse and post-stratification weighting schemes in multiple European countries. My analysis applies this method but lacks the sample size and underlying data to construct comparable analyses.

from the Marlowe-Crowne Index (Crowne and Marlowe 1960). Scales presented positive and negative personality traits, and social desirability was defined as existing when the respondent claimed socially desirable personality traits and denied socially undesirable personality traits to him- or herself (Edwards 1960 as cited in Helmes and Holden 2003). Such scales have been translated and adapted for use in other countries (Verardi et al. 2010; Vu et al. 2011). The ACV pilots use a variation of the Marlowe-Crown Index to measure tendencies toward social desirability of the adult respondent in the household. Thirteen individual questions were translated into Swahili and Nepali. An index of responses had a range of 0, reporting all socially undesirable answers, to 13, answering every question in a socially desirable way. Of the variables present in the adult survey in the ACV project, the social desirability index (SDI) score is likely to be associated with nonresponse. Agreeing to participate in t survey is a socially desirable action in itself (Harling et al. 2017; Johnson and van de Vijver 2003; Gosen 2014). Thus, respondents may have higher SDI scores than non-respondents, for whom we do have any data. I calculate the mean score of the SDI for each pilot without weights and then apply each of the different nonresponse weights to see if the mean score varies due to the application of weights[36].

Figure 2.10 shows the mean score on the SDI plus and minus the standard deviation for respondents in the three pilots (the red bar represents rural Tanzania, the blue bar represents Nepal, and the black bar represents urban Tanzania). The first column shows the raw, unweighted scores. In all three pilots, the average scores ranged between 8 and 9 with respondents in rural Tanzanian scoring the lowest (or least socially desirable) and urban Tanzania scoring highest (most socially desirable) on average.

The following three columns show the SDI scores when applying the three versions of the nonresponse weights. I conducted several sensitivity analyses using the three nonresponse weights that include boarding school children ($W_{NR1BS,}$ $W_{NR3BS,}$ W_{NR5BS}). Additionally, I tested the six final weights (W) – nonresponse weights (with and without boarding school children) combined with the base weights – presented in the section above (with the base weight three being using in the Nepal pilot as the preferred base weight). The results for these nine weights are nearly identical to the results from the nonresponse weights (without boarding school children) and are not included in the figure.

[36] I apply only the nonresponse weights to correct potential nonresponse bias in this analysis as the base weights would be the same for all households in the same primary sampling unit (PSU).

There are virtually no differences between the SDI scores when applying the three different nonresponse weights calculated from the three different response rates. There is no statistically significant difference in mean SDI score when using any of the weights compared to not using weights. In rural Tanzania, the standard deviation of weighted SDI scores is larger than the unweighted scores. This suggests that using nonresponse weights to look at SDI scores does not benefit the user of the ACV data. In some cases, weights might even increase the standard deviation.

2.8. Population Estimation from ACV Paradata

A third way to quantify potential nonresponse error, a potential source of error in the Total Survey Framework found at the transition between the sample and the respondents, is to construct population estimates from the paradata documentation of the sampling and data collection process. The sample sizes, response rates, and overall data about eligible and ineligible households in the paradata can be used to construct population estimates that can be compared to population figures found in other sources of data. The idea of estimating the population from the sample survey data presents the ultimate paradox. The sample is created as a representation of the target population because the sample was created from the target population. Therefore, how can the sample tell us anything about the population if the information about the population was unknown at the start of the sampling process? I argue that despite the ACV pilot samples being created from population data, after field work is complete, we have more information about the target population than we did prior to sampling.

The target population of the ACV pilots was households with 12-to-17-year-old residents. The sampling strategies employed in the three pilots sought to identify this subset of households from the broader populations. There are no accurate population level data sources that could provide the foundation for a sampling frame for such a specific type of household. Population figures for the rural Tanzanian village, the Arusha urban municipality, and the two municipalities in Kathmandu were only available as aggregate numbers of individuals or households. The process of sampling with the assistance of the local leaders in Tanzania and satellite imaging in Nepal was a way to create a probabilistic sample in the absence of information about the target population.

One way to assess whether the sampling and data collection represent a probabilistic sample of the target population is to compare the ACV sample paradata and calculated outcome measures and sample weights to established population statistics. In this section, I discuss the feasibility of this comparison and, when possible, I construct population estimates based on the sample data. The population estimates are constructed from paradata about the sampling process that estimated the total size of the sub-village, *mtaa*, or ward based on information from the local leaders or the process of geospatial identification of buildings. I also utilize paradata from the data collection process including identifying eligible and ineligible household and estimates on household sizes[37]. I compare my estimates to the 2012 Tanzania Population Census and the 2011 Nepal Population Census from the IPUMS-International census microdata[38]. This analysis is aimed at validating the paradata produced and collected about the sampling and data collection process by the local government officials and hired sampling teams.

2.8.1 Rural Tanzania

The pilot village in rural Tanzania was small. Posted on the wall of the village office was a table that stated there were 3,800 individuals living in 1,259 households in the village in 2015. We calculated the proportional stratified sampling methodology on the proportion of households in each sub-village from the 2015 sample. Based on the paradata from the household lists provided by the sub-village leaders, I estimate that the village had a total population of just over 3,300 individuals in 878 households during the field work period in July 2018. The estimate of the population for the village comes from the number of households that were counted on the household lists provided by the sub-village leaders. In all seven sub-villages, we were able to access the entire sub-village list and the total number of households and individuals was verified by two members of the field team[39]. Despite the three-year difference, the similarity in these numbers is reassuring.

In addition to checking the population estimates based on the posted population figures, I compare the proportion of eligible households in the village to estimates of the proportion of

[37] I do not apply the base weights constructed in section 7 in these population estimates as the base weights only represent the target populations of households with 12-to-17-year-old resident and it would be inaccurate to apply these weights to the ineligible populations.
[38] Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis, MN: IPUMS, 2019. https://doi.org/10.18128/D020.V7.2 and the Census Bureau of Statistics of Nepal and National Bureau of Statistics of Tanzania.
[39] In one sub-village, the two counts varied by approximately 20 households and we were not able to revisit the list to count a third time. In the analysis, the average of the two counts was used.

eligible households produced by the 2012 Tanzanian census. The smallest geographic unit in the IPUMSI microdata in Tanzania is the district. The village in the pilot is located in Monduli District (total 2012 population 158,929). It is not possible to identify individuals or households from the ACV pilot village in the 2012 census, but I compare the estimates for all of Monduli to the pilot village.

Using the census data, I estimate that 41 percent of households in Monduli District would have been eligible for the ACV pilot[40]. Based on the population counts in the village, I estimate that 29 percent of the total households on the sub-village lists would have been eligible for the sample. One reason for the 12 percent difference between the district as a whole and the village could be changes in the population between 2012 and 2018, such as changes in fertility and family size or migration of families out of the district. Second, it is possible that other areas in Moduli are very different from the specific village in the ACV pilot.

A final, and most likely, reason for the difference comes from the ACV population estimates. In three of the seven sub-villages, we did not record the total number of eligible households. In these three sub-villages, I estimated the number of eligible households using the proportion of eligible households from the total number of households of the four known sub-villages. The average proportion among the four known sub-villages is 29 percent (minimum = 23 percent; maximum = 39 percent). Therefore, if the three unknown sub-villages actually looked more like the sub-village with the maximum 39 percent eligibility of the total households, then the overall proportion of eligible households in the village would be closer to the 41 percent estimation found in the 2012 census data.

The incompleteness in the paradata for the rural Tanzania pilot results in discrepancies between the 2012 census data and the sample. The small size of the village also makes comparative population estimates impossible because most nationally representative datasets will rarely provide data with individual household detail at such a small geography due to confidentiality concerns.

2.8.2 Urban Tanzania

Data for the sampling process in urban Tanzania was provided by local government officials in the Arusha district office. These population figures were obtained by our Tanzanian

[40] The IPUMS-International census data for Tanzania is a 10 percent sample. Thus, statistics using weights are estimates of the total population and not actual population counts.

partners after numerous visits to the offices. The local government data was based on the 2012 census but provide aggregate population figures (by sex) for a much smaller geographic unit than the IPUMS-International microdata. IPUMS-International microdata for Arusha District can be used to compare the ACV population estimates for the specific target population of households with 12-to-17-year-old residents, similar to the rural village. The local government aggregate data can be used to assess the quality of the ACV population estimates for the entire district.

As described in section 3.2, there are ambiguities between the exact municipality boundaries in Arusha District[41]. In the 2012 census data, an estimated 42 percent of households in Arusha Urban District would have been eligible for the ACV pilot. This exactly matches my calculation that 42 percent of households in the selected cell of the selected *mtaa* were eligible. This comparison offers some evidence of validity in the sampling process in the ACV urban Tanzanian pilot.

We obtained actual *mitaa* population estimates by sex from the local government offices. With these numbers, we are not able to construct anything about household composition or eligibility, but through several assumptions, we can create population estimates at the *mitaa* level.

The sampling process in urban Tanzania, outlined in Figure 2.3, included three levels of randomization: *mtaa*, cell, and household. At each level, the paradata recorded estimated numbers of cells per *mtaa* and households per cell. I estimate the overall population size of the *mtaa* based on this paradata and assumptions about the average household size based on the collected data. I compare my estimate of the *mtaa* population size to the provided population estimates from the local government office. The local government data includes only total population sizes, but no information about how these individuals are grouped into households or the ages of the population.

The sampling team worked with each *mtaa* leader in each of the 20 selected *mitaa* to identify all of the cells in the *mtaa* and provide an estimate of the size of the cell. Cells are traditionally ten households but in the urban areas they can vary in size; if the *mtaa* leader did not know the exact number of households in the cells, we asked the sampler to probe for the leader's best estimate of size. The 20 selected *mitaa* had 7.6 cells on average (range: 1-17) with an

[41] In the ACV pilot, we sampled from Arusha City (Arusha Mjini) but the 2012 Tanzania census microdata indicates residences within Arusha Urban District (Arusha Vijijini). These two terms do not identically define the same geographic area. Prior to the start of the sampling process, our team compared which wards were located in Arusha City and which were in Arusha Urban District and found considerable overlap in the vast majority of the wards.

average of 67 households per cell (range: 25-150). In each *mtaa*, we sampled one cell. In this sampled cell, we asked the cell leader to further verify the number of households and estimate the number of people living in the cell.

Based on this information collected during the sampling process, I constructed estimates at the *mitaa* level based on the following assumptions. First, that the data provided by the *mtaa* and cell leaders is accurate and complete. The first assumption is flawed as 7 of the 20 *mitaa* do not have estimates for the size of all of the cells reported by the *mtaa* leader. This leads to the second assumption that the cells within a *mtaa* are homogenous in size so that the average number of known households per cell can be applied across all cells within a *mtaa*. This allows me to estimate the number of households in cells that do not have complete data based on the average cell size (number of households) from known cells in order to produce estimates of the total households in each *mtaa*.

The population estimates I calculate based on these assumptions end up underestimating numbers of households in each *mitaa* by approximately 50 percent of the published figures. The average difference is an underestimate of 452 households. This average is highly skewed by a single ward (Ward #2) where the published number is 4,289 households, but my calculations estimate 720 households. If I exclude this outlier ward, my estimates of the number of households per *mitaa* are underestimated by an average of 279 households.

I constructed estimates of the population of the *mitaa* based on the average number of people living in each household based on the paradata gathered from the cell and *mitaa* leaders. Similar to the estimates above, the paradata estimates of population numbers were incomplete and I made assumptions that the known average number of persons per household reported by the cell leader can be applied to all households in all cells in the *mtaa*. The estimates of population size were also underestimated by 32 percent on average, compared to the official figures from 2012.

Both the calculation of households and population estimates rely on faulty assumptions that cell sizes are homogenous, and the average cell size and household size can be applied to all cells in a *mtaa*. It is also possible that the composition of the city has changed between 2012 and 2018. Certain *mitaa* may be smaller if migrants are moving to other areas of the city or peri-urban suburbs. However, the city of Arusha is expanding rapidly, and it may be difficult for a cell leader to know what is happening among each family across the cell. Traditionally, the cells would established to represent 10 households, but increases in rural-urban migration have expanded the

population of Arusha. Even if local leaders were aware of the results published in the 2012 census, they may still underestimate exactly the number of households in their *mitaa* or cell[42].

Given these caveats, the construction of population estimates of the *mitaa* in Arusha City from the ACV paradata in urban Tanzania is not recommended. The small sample size and limited paradata about unsampled cells do not have enough accurate detail support accurate population estimates. This does not suggest that our results in the ACV pilot are not valid or inaccurate. The multi-stage probabilistic sample was implemented to generalize to the broader population of Arusha City's 12-to-17-year-old children; the population estimates for the total population was not the purpose of aim of the study, but merely a post-survey experimental analysis of the collected paradata.

2.8.3 Nepal

In Nepal, I used the estimated household and population sizes from the sampling frame complied by the sampling team through the process of identifying buildings and eligible households to estimate a total population size across all of the sampled wards in the two municipalities. The paradata collected by the samplers included estimates of the number of households in building visited and the average household sizes for eligible and ineligible households. I used these numbers to extrapolate estimates of the number of people and number of households in all of the buildings in the wards identified from the satellite images. Similar to the Tanzania pilots, I obtained aggregate population data for the wards provided by the municipality offices based on the 2011 Nepali Census. However, unlike Tanzania, the population figures do not group people into households.

In order to calculate population estimates for each sampled ward in the two municipalities, I make several assumptions similar to those made in urban Tanzania. First, I assume the data collected by the sampler is accurate and complete. Second, that the estimates of the number of households per building visited is representative across the ward. Third, that the number of people in each known household is, on average, the same across the households that were not visited in the ward. Finally, it is unknown if the buildings identified by the satellite image are residential homes. Obviously labeled buildings such as schools, hospitals, or offices in

[42] I tested the theory that cell leaders would better estimate the number of households in the cell than the *mtaa* leader by applying the number of households in the sampled cell to the other cells in the *mtaa*. The estimates were essentially the same for all *mitaa*.

OpenStreetMaps were avoided, but the analysis of population estimates assumes that the buildings identified in OpenStreetMaps were residential.

I compare my estimates of total population size to the published aggregate figures. The average published ward size is approximately 7,000 people in both municipalities. In one municipality, I overestimate the population size by 65 persons on average across the 5 wards, though the estimates for the five individual wards had a standard deviation of 1,200 people. Our local collaborators report that this area is a well-established suburban area of Kathmandu and has been residential for several decades. The stability of the neighborhood and the largely residential areas contribute to the accuracy of the population estimates compare to the published results.

In the other municipality, my population estimates are overestimates by almost 5,000 persons on average across the 5 wards (s.d.= 4,200). Based on local knowledge of the area, this second municipality is a growing area. I saw many multi-unit apartment buildings newly constructed or under construction in this municipality during the field work. The published figures from 2011 are likely to be outdated and not representative of the current population.

From the 2011 census, 34 percent of the households included a child age 12-17 across the entire Kathmandu district (which includes 11 municipalities including the 2 sampled). In the ACV pilot, approximately 23 percent and 25 percent of the households identified in the sampling process were eligible in the two municipalities in which we worked. The discrepancies in the ACV estimate and the census figures could be the difference in geographic units or the nine years between the 2011 census and the 2019 ACV data collection. Alternative explanations include the high rates of migration to peri-urban areas of Kathmandu may be predominantly adults looking for work (Graner 2001); other areas of Kathmandu may have different family compositions.

2.9 Recommendations and Conclusions

Survey data collection in developing countries is a complicated and difficult process. Researchers with limited budgets, small teams, and little formal training face challenges at every stage of the survey process. The potential coverage, sampling, and nonresponse errors found in the Total Survey Error framework are amplified in small size projects due to daily decisions made by individuals on the project. Discussing potential errors is disincentivized in publications, thus early-career researchers and students have few applied examples of how to design, collect, and

reflect on the quality and shortcomings of data. This paper aims to provide guidance to researchers who find themselves designing research that falls into a gap in the survey literature.

The ACV pilots exemplify the gap in the literature between small- and large-scale survey designs. The sampling processes of the ACV pilots mimic large-scale surveys, but the resulting sample size of the pilots is small. In large samples, individual decisions of samplers and field workers that occur randomly often do not impact a sample systematically. Therefore, a large-scale survey can reduce coverage error, sampling error, and nonresponse error throughout the establishment of a representative and probabilistic large sample. The broad representation, either in geography or population, eases the process of creating sample weights and population estimates that compare to existing source data and sampling frames. The rise of technology such as GIS imaging for sampling and tablets using CAPI (computer-aided personal interviewing) software for interviewing in survey work allows teams to identify and correct quality control issues faster than traditional methods. The standards, methods, and infrastructure created by large-scale survey organizations can be applied and adapted for use by small studies, such as ACV.

The ACV pilots also demonstrate the practical, social, and logistical challenges of establishing a sampling frame of atypical households—those in which at least one 12-to-17-year-old child resides. Most large-scale household surveys seek to interview adults; therefore, the literature assumes that probabilistic samples of households will yield an adult respondent. Literatures on sampling "hard to find" or hidden populations other methods of sampling including snowball sampling or convenience sampling (Watters and Biernacki 1989; Salganik 2006; Magnani et al. 2005; Sadler et al. 2010). Potential coverage and sampling errors that arise during the sampling process of a specified target population can also amplify potential nonresponse errors demonstrated in the process of creating outcome measures such as response rates, sample weights, and population estimates.

Based on the descriptive accounts of the ACV pilot studies in rural Tanzania, urban Tanzania, and peri-urban Nepal, I offer the following recommendations for adapting and applying sampling methods and data collection techniques found in text books and literature based on large-scale survey collection in developing countries[43]. Many of the recommendations here do not contradict the teachings from the field of survey research but instead highlight specific areas that may be more relevant for small-scale studies. I organize these recommendations by the two areas where the ACV pilots experienced the most challenges: social and logistical.

[43] An excellent resource is Survey Research Center (2016).

2.9.1 Social Recommendations

- Work with local survey organization partners to build relationships with key stakeholders such as local leaders, community leaders, and field team members.
- Build legitimacy with local government officials. This process may be time consuming and bureaucratic. Local government officials may be key in providing existing population figures needed to create a sample, and they may be able to assist in making key introductions to local leaders. Local leaders don't care about your research and your research is not a priority for them.
- Establish training and field work protocols for all members of the sampling and data collection teams. Work with local partners to learn and incorporate norms and cultures of survey collection into the training and protocols.
- Hire field team members that are trained in human-subjects and ethical research practices. Field team members should be respectful of the ethical protocols while also friendly and efficient when interacting when interacting with participants.
- The members of the sampling or pre-sampling screening procedure team may not be the same individuals as being hired to do the data collection. The sampling process may require a different set of skills than the data collection.

2.9.2 Logistical Recommendations

- Be creative about sources of data to construct sample frames. Identifying information about the eligible and ineligible populations as well as the sampled and unsampled populations can aid in the production of sample weights and population estimates to analyze the validity of the data collection and sampling processes.
- Budget enough money and time to properly conduct the sample. Take into consideration the travel time between locations and the time spent in each location.
- Whether you have a complex or simple sampling design, clearly communicate to the sampling and data collection team the information to be collected. Systematize the collection of paradata and documentation of the daily process.
- If possible, use digital technology such as CAPI to track paradata in the sampling and data collection process. Some survey software includes features in the program to make this easier. This includes interactions with eligible and ineligible households, results of each contact with a household, and comments about each household.

- Over-document everything. One benefit of a small-scale study is that the sample size may be small enough to have a discussion about each household individually. Daily debriefing notes provide justification and clarification for categorization of interactions with households.

These social and logistic recommendations are particularly important for small-scale studies that aim to make generalizable claims about the validity and statistical accuracy of the survey results. The general techniques of identifying and correcting nonresponse error—outcome measures such as response rates, sample weights, and population estimates—are data intense calculations that require additional data and paradata beyond the results of the survey. To assess nonresponse error, it is not enough to know about the participants included in the survey, researchers must know about the households that did not respond, were not identified, and were not eligible.

It is important for researchers to properly specify definitions of outcome measures reported. As discussed in section 2.6, outcome measures can be subtly manipulated in a way that is not incorrect or unethical but allows researchers to portray the quality of their data in different ways. I therefore recommend researchers follow the AAPOR recommendations of proper definitions and clarification in all publications and reports. This does not suggest that researchers should not be strategic in reporting measures that accurately reflect the sampling and data collection process, but overall transparency will strengthen the validity and reliability of the data and results.

It may be that the results of the analysis of nonresponse error do not contribute to the overall interpretation of the survey results. In the ACV pilots, the high response rates and low refusal rates suggest the interpretation of the survey results may not be influenced by nonresponse bias. The accurate comparison of the proportion of the total population eligible in existing census data compared to collected ACV paradata about eligible and ineligible populations suggest minimized coverage and sampling errors in the ACV pilots.

On the other hand, the calculated base weights (weights designed to inflate the sampled population to the target population), and population estimates of the target population were less accurate. Specifically, the urban Tanzania and peri-urban Nepal pilots were designed to be generalizable to large populations and geographies[44]. The team collected paradata about eligible

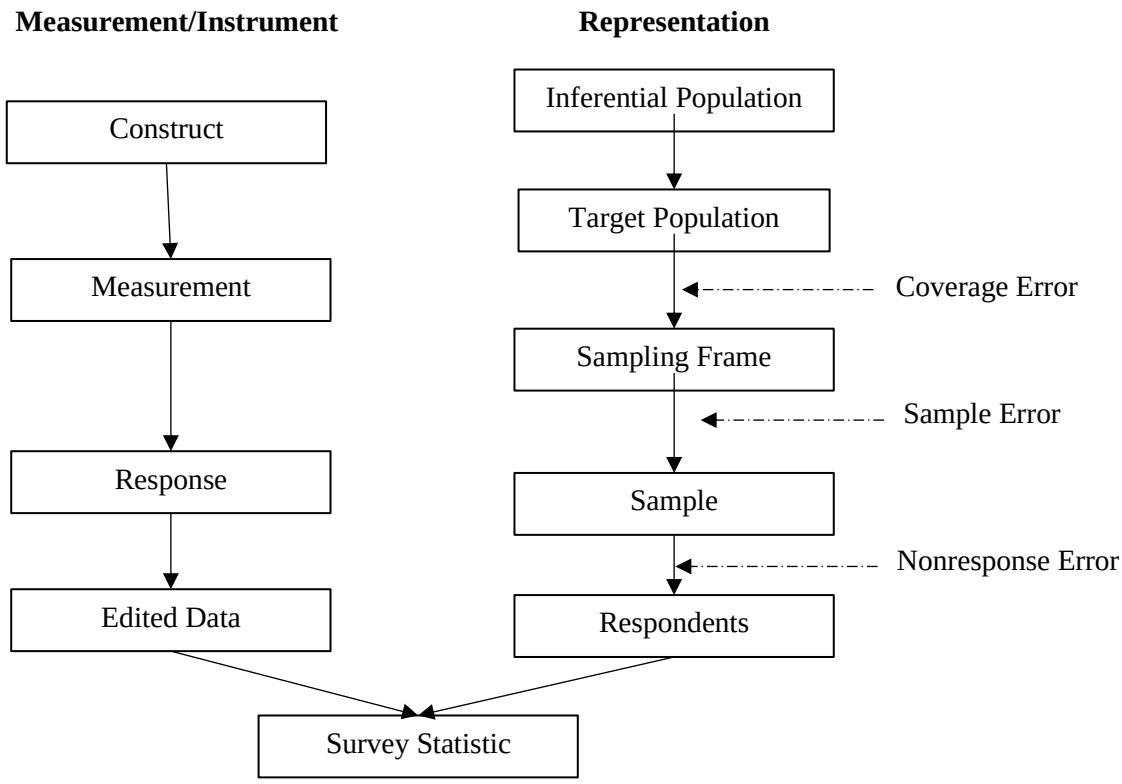
[44] The small size of the rural Tanzania village did allow for the calculation of sample weights and population estimates that had smaller variation than in the other pilots. But we are still limited by assumptions and outdated data provided by local government officials.

households in the sampled communities. But the paradata required too many assumptions about unsampled and ineligible households and populations to produce accurate base weights and population estimates. This does not invalidate the ACV pilots in these two locations—the sampling process was probabilistic and representative. But it instead suggests the limitations of producing such estimates from small-scale surveys. These calculations are data intensive and require significant social and logistic effort that was beyond the scope, budget, and timeframe of the ACV pilots.

The reality of data collection is messy. Social and logistic challenges are faced by everyone doing social science research and the challenges are unique to every pilot location. Being transparent about potential error in survey processes could jeopardize publication and overall validity of survey results. But it is the only way that readers of academic literature can communicate areas for growth. Sampling designs do not have to be complicated, but simple designs need to be executed properly. It is difficult to use statistics to adjust away errors that occurred during sloppy data collection.

2.10 Figures

Figure 2.1: Total Survey Error Components Linked to Steps in the Measurement and Representational Inference Process



Adaptation of Figure 3 in Groves & Lyberg (2010)

Figure 2.2: Flow Chart of Sampling Process in Rural Tanzania

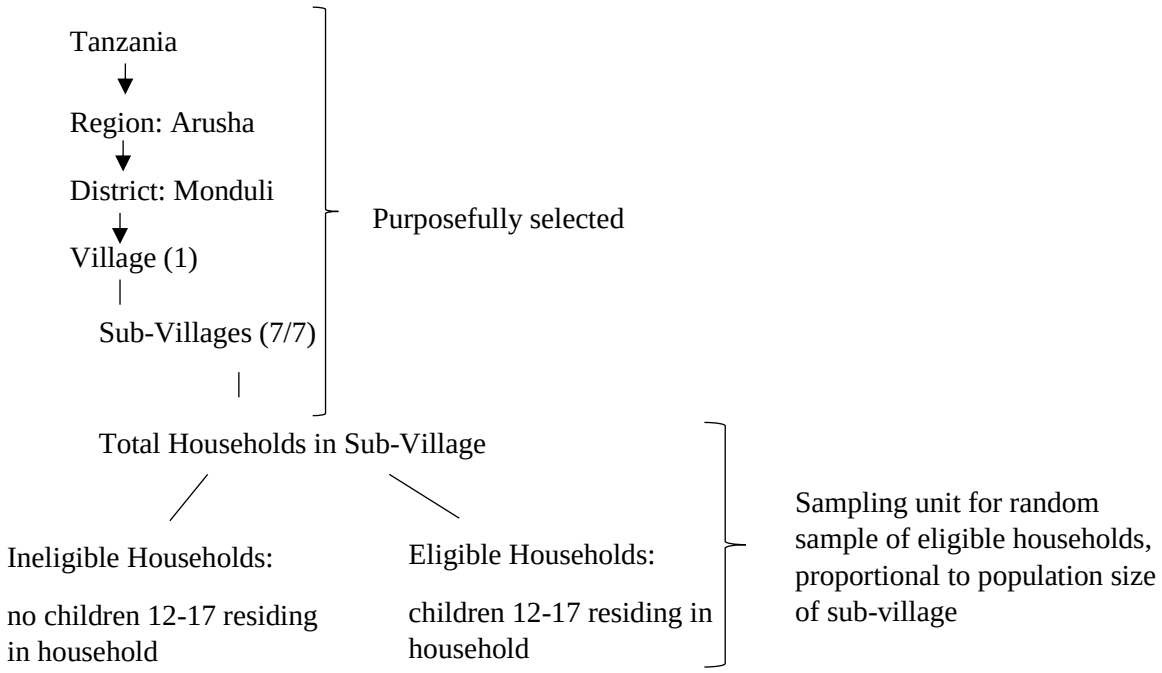


Figure 2.3: Flow Chart of Sampling Process in Urban Tanzania

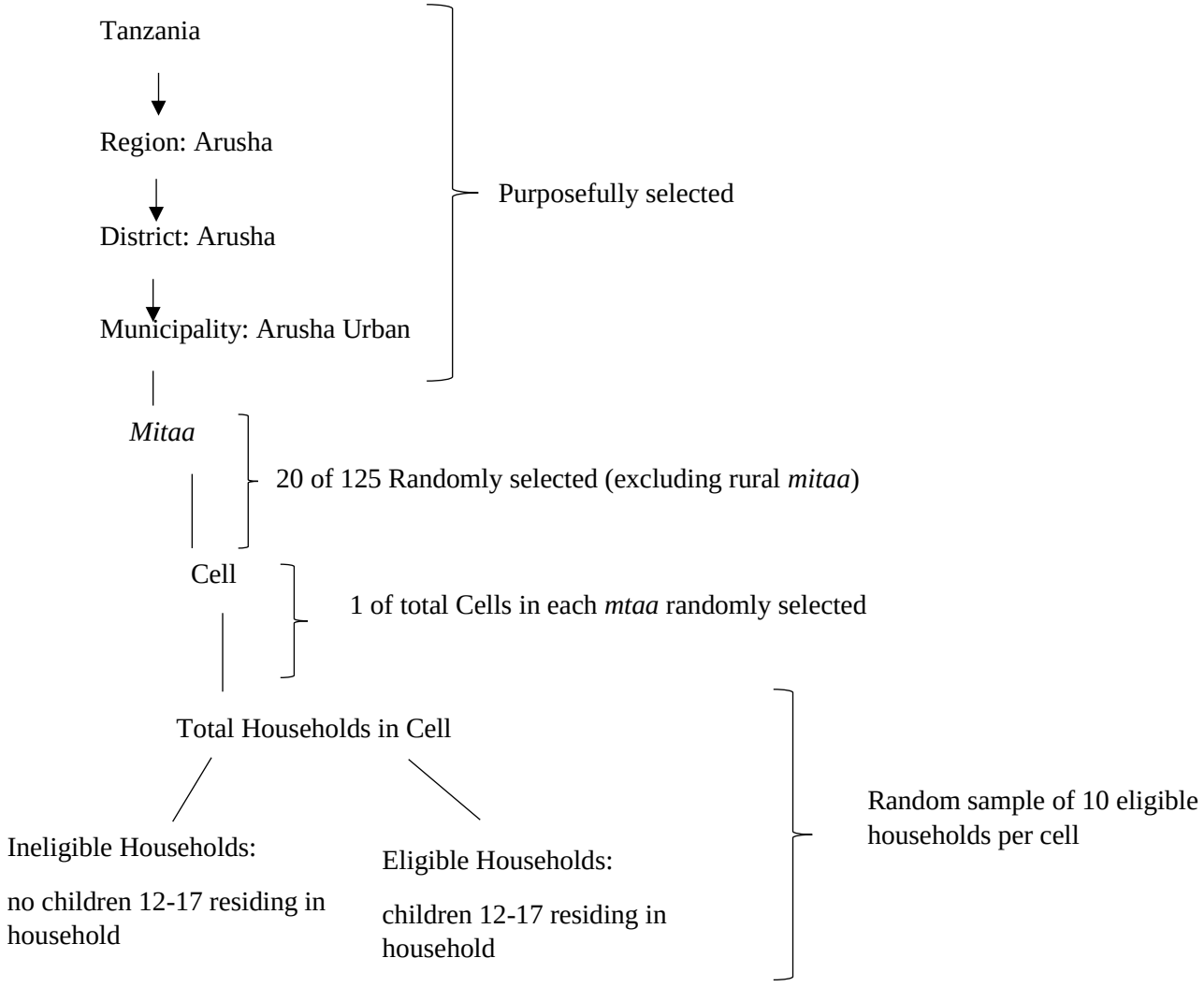


Figure 2.4: Flow Chart of Sampling Process in Nepal

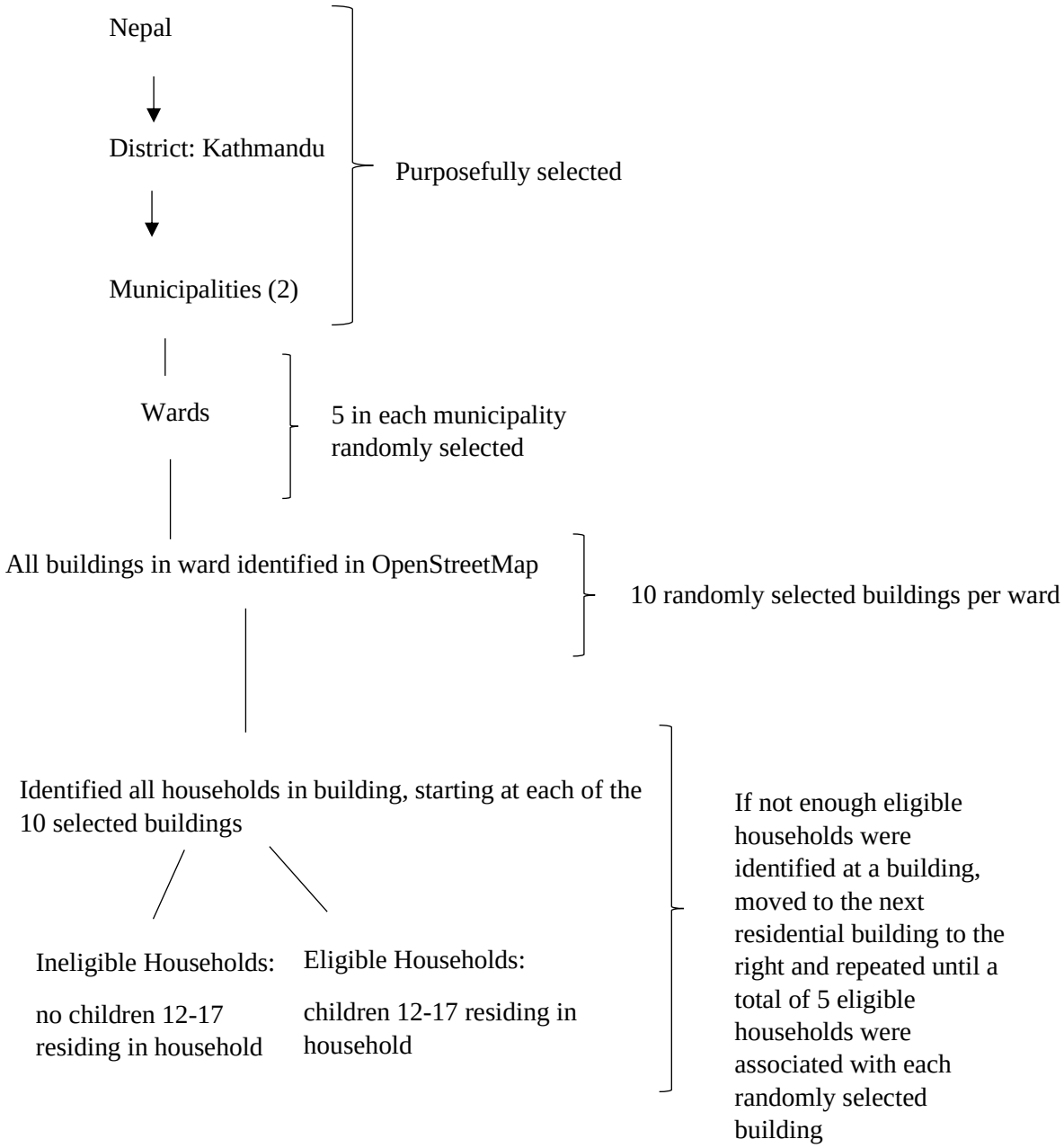
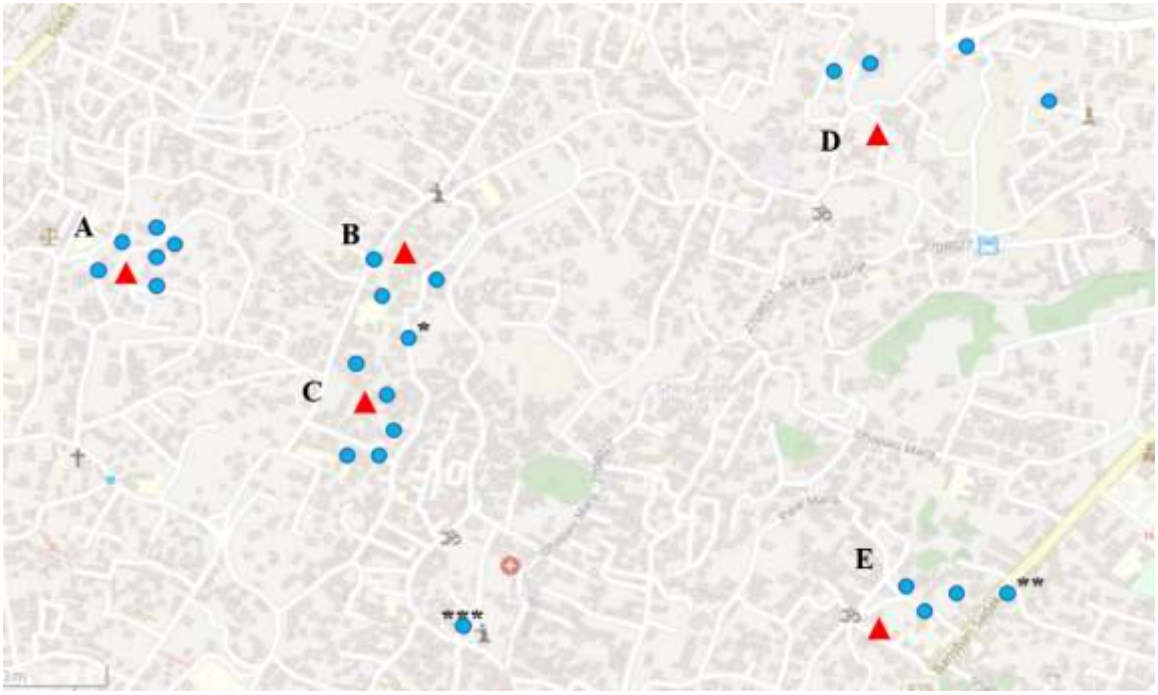


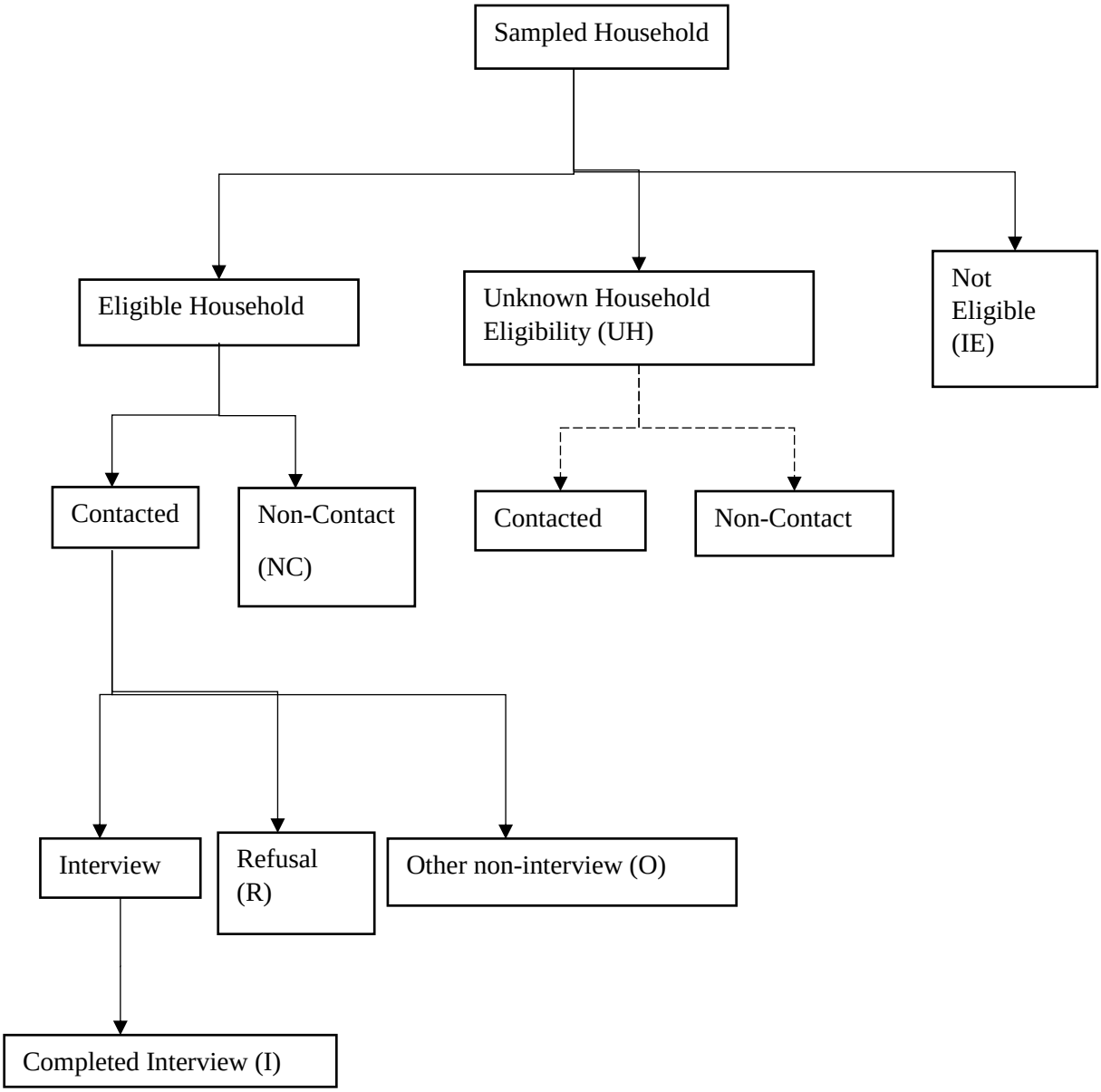
Figure 2.5: Stylized Map to Show Geospatial Sampling in Peri-Urban Nepal

Red triangle represents sampled buildings
Blue circles identify Eligible households
*Represents a household that falls between 2 sampled buildings
**Represents a household outside of the ward boundary (on the other side of the yellow road on the map)
***Represents a household interviewed in a location far from any originally sampled household, such a community center or café.



Note: Figure does not represent actual geographic locations of sampled buildings or interviewed respondents

Figure 2.6: Flow Chart of AAPOR Categorizations of Household Eligibility



Adapted from Beerteen, Lynn, Laiho, and Martin (2015)

Figure 2.7: Rural Tanzania Final Weights (W) by Sub-village (1-7)

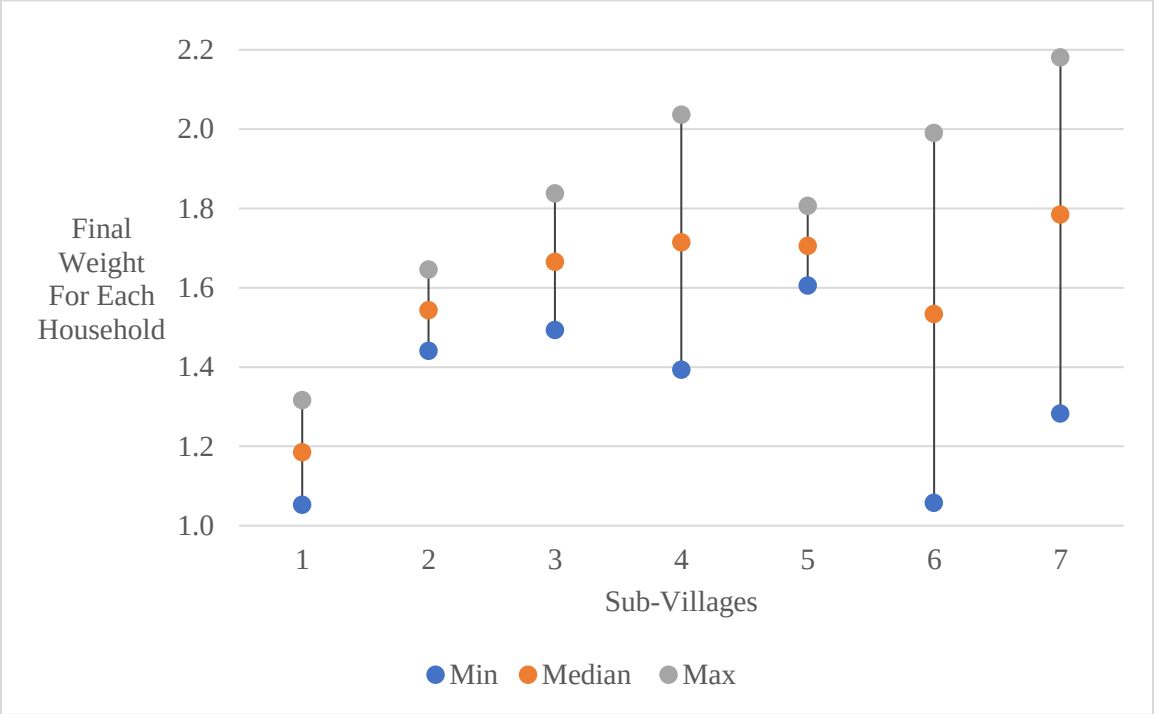


Figure 2.8: Urban Tanzania Final Weight (W) by Mtaa

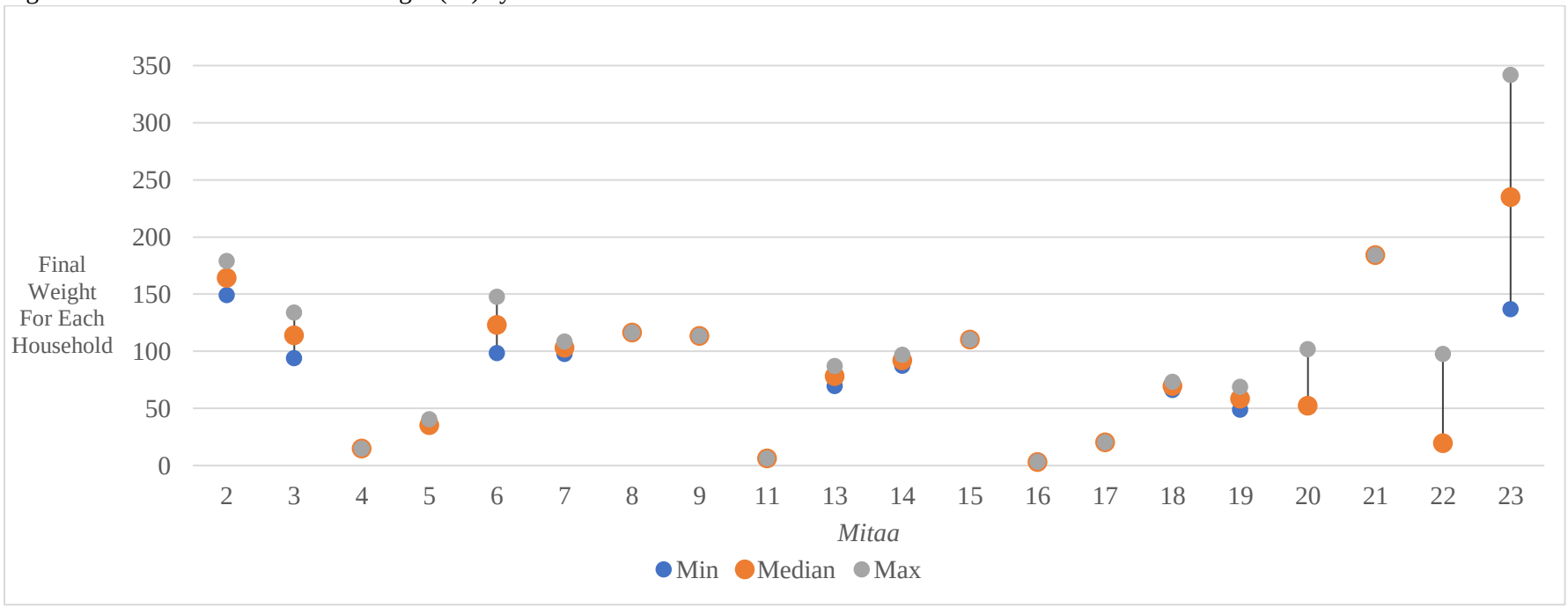
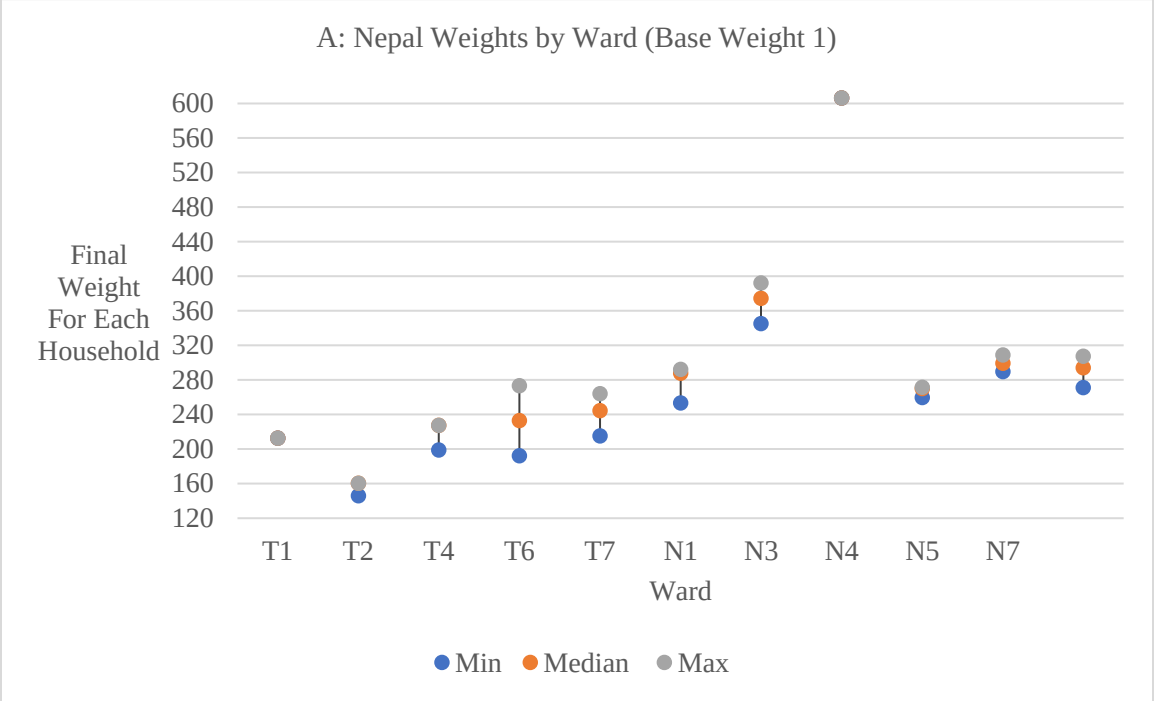


Figure 2.9 A & B: Nepal Final Weights by Ward (10 wards)
Note the difference in magnitude along the y-axis between Figure A and Figure B.



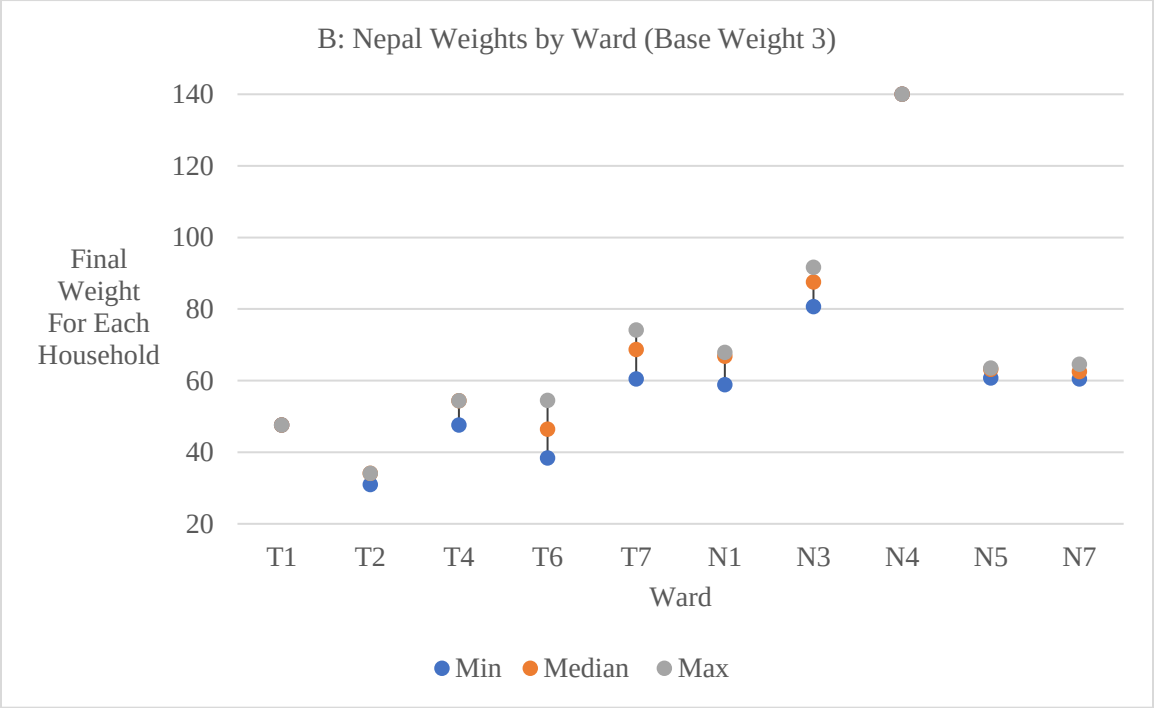
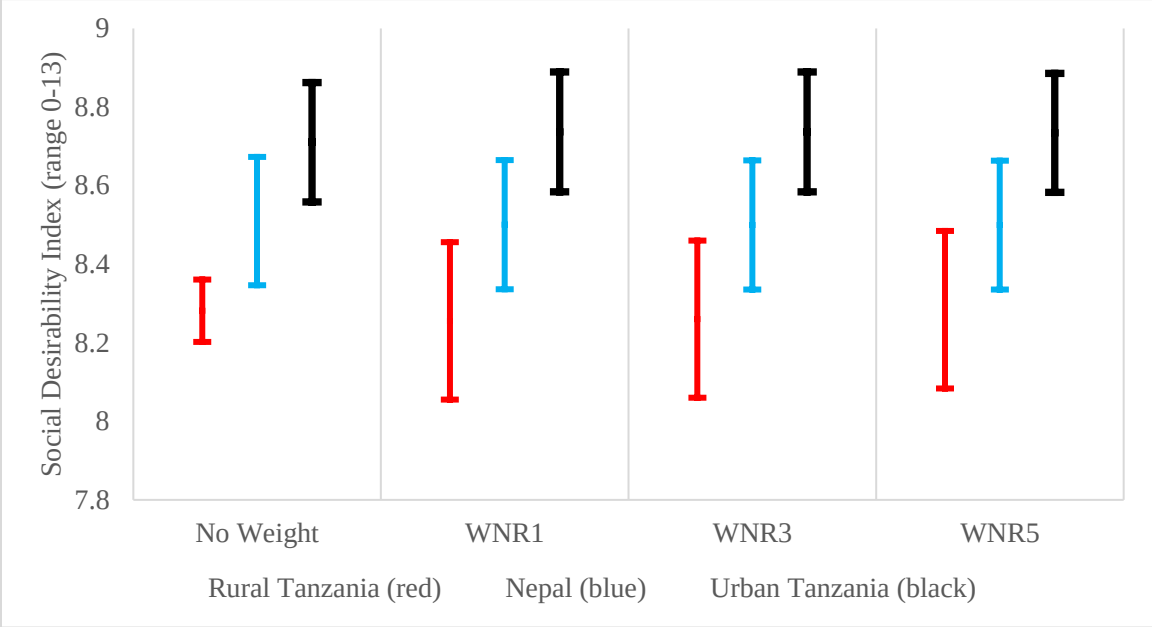


Figure 2.10: Comparison of Nonresponse Weights on the Social Desirability Index (SDI)



2.11 Tables

Table 2.1: Summary of sampling methods as proposed in Fottrel and Byass (2008) and the application of the technique to the context of the 3 pilot studies

Pilot	Sampling Method	Technique	Translation to Context
Rural Tanzania	Proportional Stratified Sampling	Step 1: Determine the proportion of sampling units needed in each strata	Among 7 sub-villages of village, determine number of households needed based on posted population sizes
		Step 2: Assign a random number to each sampling unit	Assign households a number based on order
		Step 3: Select sampling units from each strata using simple random methods until the desired sample size and ratio between strata is obtained.	Using a random number generator, select households with children 12-17 from each of the sub-villages until the determined sample size for each sub-village is reached.
Urban Tanzania	Multi-Stage Sampling	Step 1: Randomly select geographical area for sampling	Randomly selected 20 of 125 *mitaa* in predetermined urban city
		Step 2: Assign a random number to each sampling unit in the select area	Within selected *mitaa*, identified all cells and randomly identify one (or more)
		Step 3: Sort sampling units by their random number	Within the selected cell, identify all households with children 12-17
		Step 4: Select sample units in ascending order of random number until desired sample size is reached	Randomize order of households with 12-17 and sample until desired sample size within each *mtaa*
Urban Nepal	Geographically Dispersed	Step 1: Randomly select # geographic areas	Purposefully selected 2 municipalities within Kathmandu District. ACV did not do this randomly to maximize variation within the municipalities.
		Step 2: Assign a random number to each sampling unit in each of the selected areas	Identify wards within municipalities
		Step 3: Sort sampling units by their random number	Randomly select 50% of the wards in each municipality
		Step 4: Select sampling units in ascending order of random number until 50% of the desired sample in selected from each geographic area	Identify all buildings (sampling units) within selected wards. Randomize order of sampling units. Establish 10 as the target number of units per ward. Select first 10 random sampling units within each ward. Further steps identify which households contain children 12-17 using a random walk method.

Table 2.2: AAPOR Outcome Rate Formulas (AAPOR 2016)

Response Rates	
RR1	$\frac{I}{(\mathrm{I}+\mathrm{P})+(\mathrm{R}+\mathrm{NC}+\mathrm{O})+(\mathrm{UH}+\mathrm{UO})}$
RR3	$\frac{I}{(\mathrm{I}+\mathrm{P})+(\mathrm{R}+\mathrm{NC}+\mathrm{O})+\mathrm{e}(\mathrm{UH}+\mathrm{UO})}$
RR5	$\frac{I}{(\mathrm{I}+\mathrm{P})+(\mathrm{R}+\mathrm{NC}+\mathrm{O})}$
Cooperation Rate	
COOP1	$\frac{I}{(\mathrm{I}+\mathrm{P})+\mathrm{R}+\mathrm{O}}$
COOP3	$\frac{I}{(\mathrm{I}+\mathrm{P})+\mathrm{R}}$
Refusal Rates	
REF1	$\frac{R}{(\mathrm{I}+\mathrm{P})+(\mathrm{R}+\mathrm{NC}+\mathrm{O})+(\mathrm{UH}+\mathrm{UO})}$
REF2	$\frac{R}{(\mathrm{I}+\mathrm{P})+(\mathrm{R}+\mathrm{NC}+\mathrm{O})+\mathrm{e}(\mathrm{UH}+\mathrm{UO})}$
REF3	$\frac{R}{(\mathrm{I}+\mathrm{P})+(\mathrm{R}+\mathrm{NC}+\mathrm{O})}$
Contact Rates	
CON1	$\frac{(I+P)+R+O}{(\mathrm{I}+\mathrm{P})+(\mathrm{R}+\mathrm{NC}+\mathrm{O})+(\mathrm{UH}+\mathrm{UO})}$
CON2	$\frac{(I+P)+R+O}{(\mathrm{I}+\mathrm{P})+(\mathrm{R}+\mathrm{NC}+\mathrm{O})+\mathrm{e}(\mathrm{UH}+\mathrm{UO})}$
CON3	$\frac{(I+P)+R+O}{(\mathrm{I}+\mathrm{P})+(\mathrm{R}+\mathrm{NC}+\mathrm{O})}$

Key:

- Eligible Households that were interviewed
  - Completed Interview (I)
  - Partial Interview (P)
- Eligible households that were not interviewed
  - Refusals and break-off (R)
  - Non-contact (NC)
  - Not interviewed for other reasons (O)
- Households not interviewed and unknown if the household would be eligible (UH)
  - Unknown for any other reasons (UO)
- Ineligible households (IE)
- e = estimate of the probability of UH being eligible

Table 2.3: Specific Sub-Categorizations of Unknown Households (UH)

"Not attempted"	In these households, the field research team did not attempt to contact the household at all via phone or visit. In Nepal, the geographic layout of the sampling meant that some wards were very remote and if the team ran out of time to visit all households, the ward was not visited again.
"Household unsafe or unable to reach"	The team often encountered households that were located in areas that the jeep or transport was unable to travel to. Additionally, we encouraged field researchers not to put themselves in danger if approaching a house with a guard dog, if they were unable to call or shout to the inhabitants of the house.
"Unable to locate"	This category includes notes in the paradata like "gave up", "didn't find", or "no one home". In other words, there was an attempt to find household members, but it was unsuccessful.
"Unable to make contact via phone"	In a special category for Nepal, problems with the phone connections were recorded. In Nepal, there was a pre-screening process that collected the phone numbers of sampled households about 2-6 weeks before the field team went to the locations[45]. Because the primary contacting of the household was done via phone, the Nepal team encountered barriers when the service was shut off or the line was blocked.

[45] This is in contrast to Tanzania, where a community leader was showing the field team to the home in person.

Table 2.4: AAPOR Reporting Outcomes Measures for ACV Pilots in Rural Tanzania, Urban Tanzania, and Nepal

	Rural Tanzania	Rural Tanzania (BS)	Urban Tanzania	Urban Tanzania (BS)	Nepal	Nepal (BS)	Total	Total (BS)
Response Rates								
RR1	76%	64%	77%	73%	79%	76%	78%	72%
RR3	80%	67%	77%	73%	80%	77%	79%	73%
RR5	91%	75%	84%	80%	86%	83%	87%	79%
Cooperation Rates								
COOP1	99%	99%	94%	94%	88%	88%	93%	93%
COOP3	100%	100%	95%	95%	90%	90%	94%	94%
Refusal Rates								
REF1	0%	0%	4%	4%	8%	8%	5%	4%
REF2	0%	0%	4%	4%	9%	9%	5%	5%
REF3	0%	0%	4%	4%	9%	9%	5%	5%
Contact Rates								
CON1	77%	81%	82%	83%	90%	90%	83%	85%
CON2	80%	83%	82%	83%	91%	91%	85%	86%
CON3	90%	75%	89%	85%	96%	93%	92%	85%

BS = inclusion of households with eligible children away at boarding school students as non-contact (NC). Definitions of AAPOR outcome measures found in Table 2.2

Chapter 3

Siblings of Children with Disabilities in Tanzania: Prevalence and Education Outcomes

Abstract:

Sibling relationships significantly shape the experience of childhood and future life outcomes, particularly if one of the siblings has a disability. There is a significant gap in the literature regarding persons with a sibling with a disability in developing countries. In this chapter, I establish nationally-representative estimates of the number of children living with siblings who have disabilities in Tanzania using the 2012 Census. Existing theories of sibling relationships based in Western or Euro-centric cultures suggest that siblings of persons with disabilities may experience different educational outcomes than persons without siblings with disabilities due to increased caretaking responsibilities and/or family dynamics. I show that application of such theories in Tanzania yields statistically insignificant results on the likelihood of school enrollment for Tanzanian children with siblings with disabilities compared to children who do not have siblings with disabilities. These null results suggest the value of considering cultural differences in sibling relationships among children with and without disabilities in future research in developing countries to avoid making assumptions about sibling relationships based on Western understandings.

3.1 Introduction

> "Siblings always *matter*. How siblings should relate to each other, what to call them, and what resources they are to have and share is important to all cultures."
>
> (original emphasis, Weisner 1989, p.g. 14)

Brothers and sisters are ubiquitous, but the experience of such relationships varies across cultures. The roles and expectations of siblings in childhood and adulthood, how siblings interact with each other, and how siblings are situated within the greater context of the family all differ, given societal norms. Some scholars reduce differences between cultures to distinction between *industrialized* countries – those in North America and Europe, for example – compared to all

other (*non-industrialized*) countries. Sibling relationships in industrialized cultures are characterized as supportive and positive interactions that benefit socialization and emotional wellbeing. Regarding nonindustrial societies, discussions emphasize caregiving labor and the essential role of siblings in household functioning (Cicirelli 1994). While the language of industrialized and non-industrialized countries is frequently modernized as *developed* and *developing* countries, this narrow, Euro-centric, and outdated binary hides the rich variety of experiences not only *between* cultures but also *within* cultures (Weisner 1993; McGuire and Shanahan 2006).

One variation on a sibling relationship that occurs within all cultures is the relationship between siblings when one or more of the siblings has a disability. Just as described above, siblings of persons with disabilities play important roles in the support and socialization of their brothers and sisters, responsibilities that change across the life course, and they navigate these relationships within their broader family interactions. Yet, children and adults who have siblings with disabilities do experience childhood and sibling relationships differently from peers from the same cultural background who do not have siblings with disabilities (Priestly in Sanchez 2016; Stoneman 2005).

Much of the literature on siblings of persons with disabilities focuses on childhood development and psychological effects of the relationship on the sibling without the disability (Rossiter and Sharpe 2001). Psychology studies highlight positive attributes such as patience, warmth, and confidence while also hinting at potential long-term consequences such as (negative) mental health issues (Vermaes et al. 2012). Other research areas situate the experience of siblings of persons with disabilities in the context of family relationships, considering relationships with parents and dynamics between siblings as affected by birth order or gender (Schuntermann 2007). Some studies of families assess whether educational outcomes or test scores are lower for children with siblings with disabilities (Fletcher, Hair, and Wolfe 2012; Black et al. 2020). Finally, other studies consider changes over time in the roles of siblings with and without disabilities as children, adolescents, and adults (Doody et al. 2010; Hall and Rossetti 2018).

Such studies, described in detail below, characterize individuals with siblings with disabilities as having "outcomes" that can be isolated and compared to "outcomes" for individuals in typical sibling relationships. For example, positive personality characteristics or poor test scores can be identified as outcomes causally related to having a sibling with a disability. The very nature of describing atypical sibling relationships as "atypical" reinforces the assumption that there is a typical or even ideal type of sibling relationship that cannot be achieved between siblings when one or more of the siblings has a disability (Hastings 2016). Assumptions

about ideal sibling relationships reflect long-standing research traditions of holding white, middle-class, nuclear families from industrialized or developed countries as the ideal comparative standard (Stoneman 2005; Zukow 1989). Research on individual "outcomes" for individuals with siblings with disabilities has largely ignored cultural factors which fundamentally shape how individuals and families experience siblings and disability in ways that are different from Western-centric norms (Harry 2002). As Weisner (1993) argues, an individual sibling's disability, socio-emotional wellbeing, family circumstance, or stage in the life-course are not the most important aspects of understanding the experience and outcomes a sibling of a person with disability may have; the most important aspect to know about the siblings is the *culture* in which the siblings are growing up (paraphrased with original emphasis, pg. 51).

Culture is vital to understanding the entire picture of sibling relationships and their intersection with disability. This includes differences in the cultural expectations and norms between siblings without disabilities and siblings with disabilities that occur within all cultures, but also differences between cultures in developed and developing countries. At its very core, the concept of "disability" has a long history of multiple interpretations and identity. The most prominent are the medical model of disability and a social model of difficulty or impairment (Sanchez 2016; Insel 2013). The medical model of disability is often applied to developing countries without regard for historical or cultural nuance (Grech 2015). The ways people around the world interpret if someone has a disability, what defines as disability, what is the cause of the disability, what is the attitude towards persons with disabilities, and even language surrounding abilities vary drastically (Stone-MacDonald and Butera 2012; Nyangweso 2018; Ravindran and Myers 2012). Cultural interpretations of disability have been studied extensively through high-quality, yet small-scale ethnographic or qualitative (Chamie 1989). Such studies are able to consider contextual history and cultural. Any attempts to scale up and create large-scale demographic statistics about disability struggle with variations in categorical definitions of disability across cultures (examples include Maulik and Darmstadt 2007; Croft 2013; Fujiura, Park, and Rutkowski-Kmitta 2005).

Statistics on persons with disabilities that are collected and calculated are limited, culturally complex, and culturally defined[46]. Such limitations about data on persons with disabilities mean that quantitative or demographic information about *siblings and families* of

[46] It should be recognized that some theories in developed and developing countries resist the medical model of disability because it is founded on assumptions that label or categorize persons with disabilities in ways that may stigmatize individuals with disabilities and their families (Söder 1989; Susman 1994; Green et al. 2005).

persons with disabilities in the developing world cultural context is virtually non-existent. The experiences of siblings of persons with disabilities in developing countries have rarely been studied, even in small qualitative and ethnographic studies. When they are, there is rarely a recognition that assumptions and theories about the experience and outcomes of siblings of persons with disabilities were established within a Western context and may not be applicable to non-Western cultures. Some notable exceptions are described later.

In this chapter, I address both of these limitations with a specific cultural focus on the East African country of Tanzania. The primary objective of this chapter is to establish nationally-representative estimates of the number of children living with siblings who have disabilities using the Tanzania 2012 Population Census, the most recent census publicly available. I focus on children as is typically done in the literature on sibling relationships, because children have considerable impact on their siblings' experiences and life trajectories (Kramer and Conger 2009). As a secondary objective, I use these estimates to compare school enrollment status – a variable found in the Tanzania 2012 census – for children with and without siblings with disabilities. I pose the question: are children with siblings with disabilities less likely to be enrolled in school than their peers without siblings with disabilities?

In many ways, this analysis about school enrollment is guided by the theoretical focus on "outcomes" for siblings of persons with disabilities, as is often the case for literature of siblings in developed countries. However, while studies of educational outcomes for siblings of persons with disabilities in developed countries find negative effects, I find no significant difference in school enrollment for children with and without siblings with disabilities. These null results suggest that typically Western assumptions are not appropriate in Tanzania due to differences in definitions and interpretation of disability and the experience and expectations of sibling relationships. For future work, my results suggest that a theoretical framework based in a bioecological or sibling embedded systems framework, as described below, will be more appropriate.

Given the limitations of census data, the analysis on school enrollment is one example of the application of demographic information about siblings. Other analyses can use the methodology described to identify siblings to further expand the literature about *within* cultural differences in sibling relationships among children with and without disabilities beyond the assumptions made by Western scholars.

In this chapter, I use the term "disability" to refer to a broad interpretation of both a *medical* and *impairment*-based paradigm of disability and also the more inclusive concepts of

difficulty and *barriers* for individuals, families, and communities. I discuss the historic and cultural implications of these conceptualizations in the literature and how they relate to disability statistics and data collected in census data. I highlight excellent articles that explore the implications for the quantification of disability of using impairment-centric language compared to difficulty-centered language (for example, Loeb, Eide, and Mont 2008; Priebe 2018; Zola 1993)[47].

In some sibling literatures, a child without disabilities is referred to as a "typically developing" sibling, "non-disabled" sibling, or "non-impaired" sibling. I prioritize the use of person-first language for "persons with disabilities" (in contrast to "disabled person") and therefore do the same when I am referring to children without disabilities or siblings without disabilities. Throughout this chapter, "sibling" refers to a child who (unless otherwise specified) does not have a disability and has a brother or sister with a disability. Other literature in the field uses the phrase "sibling pairs" to refer to the relationship between two individuals: one without a disability and a corresponding to a brother or sister with disabilities. As my focus in this chapter is all siblings within a household, this language of a "pair" or "dyad" does not always apply: a family may have multiple children where one, some, or all may have disabilities. Finally, I refer to individuals having a disability while also recognizing individuals may have or be living with one disability or multiple disabilities, whether medically diagnosed or not.

3.2 Literature

3.2.1 Sibling relationships theories

Studying the relationship between siblings when at least one of them has a disability has undergone many changes since the 1980s. Much of the early research on sibling relationships contrasted an ideal sibling relationship with one in which the "typically-developing" sibling was in some way negatively impacted by their sibling's disability. Stoneman (2005) argues that, "Although this conceptualization has spawned a large body of research, it is an empty conceptual framework, yielding little in the way of usable or valid information" (page 339). Research focused on the outcomes of the sibling without a disability has historically been conducted without consideration of theoretical constructions of family dynamics (Stoneman and Berman

[47] As this chapter also touches on education and school enrollment, I note that the term "special education" frequently appears in literature about education of persons with disabilities.

1993, page 356). New theories about families and sibling relationships among individuals with and without disabilities were developed to incorporate nuanced, holistic, and dynamic sibling relationships. I highlight five theories[48]:

- Family Systems Theory
- Family Stress Theory
- Life Course Theory
- Bioecological Theory
- Sibling Embedded Systems Framework

The first three theories position the relationship between siblings within literature about family relationships. Bioecological theory and the sibling embedded systems framework both situate sibling relationships within a larger network of societal and community-based systems. The sibling embedded system framework is the only theory of the group specifically developed to conceptualize sibling relationships where at least one sibling has a disability; the other theories were adapted to the field from psychology, sociology, and family studies.

As the field of disability studies shifted from a medical view of impairments to a social view of disability, the study of sibling and family dynamics, particularly with young children with disabilities, also paralleled this development. However, as theories about disability identity, interpretation, and social construction have continued to grow and interact with other theories of family, feminism, and racial identity, similar movement in the field of sibling relationships among persons with and without disabilities has not fully incorporated such theories, at least in published literature. It is also recognized that the theories highlighted in this chapter were developed predominantly by Western theorists and researchers. The comparative absence of theory on sibling relationships among persons with and without disabilities from developing countries perspectives is inextricably linked to the lack of culturally appropriate research on siblings in developing countries, which is discussed at depth below.

Family Systems Theory

The family is a complex system of interwoven relationships that affect the outcomes and experiences of children (Cox and Paley 2003). Family systems theory situated siblings within the context of other family relationships such as the parent-child dynamic of power. Siblings of persons with disabilities are limited from fully participating in what might be called a "normal

[48] For a full summary of theories, including several that I have opted not to include, see Graff, Neely-Barnes, and Smith 2008.

sibling relationship," and they face other barriers as they assist their sibling in day-to-day activities (Priestly 2003 in Sanchez, 2016). Scholars, particularly developmental psychologists, have focused on how having a sibling with a disability might result in measurable, usually negative, long-term effects on the child such as anxiety, depression, and frustration for the child without a disability (Bischoff and Tingstrom 1991; Pilowsky et al. 2004; Ross and Cuskelly 2006; Hartling et al. 2014). However, other studies and systematic reviews of the literature suggest these negative studies neglect positive outcomes of having a sibling with a disability, including patience, leadership, and confidence in advocacy (Burke, Arnold, and Owen 2015; Hartling et al. 2014; Skotko and Levine 2006; Rossiter and Sharpe 2001; Stoneman 2001; Orsmond, Kuo, and Seltzer 2009).

Critics point out that many of these studies using a family systems theoretical framework rely on parental proxy reporting on behalf of the children and on medical indices to measure various psychological outcomes (Roper et al. 2014). Other research looks at the positive relationship "typically developing" siblings have on their siblings with disabilities (Begum and Blacher 2011). This type of research emphasizes the sibling-sibling relationship in the family systems theory as being multi-directional. Regardless of the positive or negative outcome studied, most studies that look at siblings within the field of disability studies employ the family systems theoretical framework. The focus is on the individual relationships among siblings and individual experiences, often in "pairs" or "dyads." These types of studies are usually small-scale, qualitative, purposefully sampled, and cross-sectional or retrospective.

Family Stress Theory

Family stress theory is a sub-category of family system theory; it highlights that outcomes for siblings may be the result of broader circumstances for families where a child or individual has a disability. Family stress theory suggests that the presence of a child with a disability in the family can result in negative outcomes for the family as a whole due to resource allocations, both monetary and emotional, to accommodate the disability. Children with disabilities are disproportionately in families in poverty in the USA and around the world (Fujiura and Yamaki 2000; Groce 2003). Households where one or more individuals have a disability are more likely to report lower levels of adult education and adult employment and inadequate housing according to US census data in 2000 (Shandra et al. 2012).

One main application of family stress theory to sibling relationships is in reference to "sibling spillover" effects. "Sibling spillover" suggests that there are effects, usually negative, that come from siblings with disabilities needing more attention or a disproportionate allocation

of family resources. In the United States, two recent studies show that children who have a sibling with a disability have lower test scores compared to children living in households without siblings with disabilities (Fletcher, Hair, and Wolfe 2012; Black et al. 2020). These quantitative studies control for family characteristics, but the emphasis of the studies using family stress theory is still on the educational scores of the individual children without disabilities in the sibling pairs.

Life Course Theory

A life course approach to sibling relationships emphasizes that sibling relationships are often one of the longest relationships that individuals have with another person. A growing literature on adult siblings finds complex and evolving sibling relationships between adult siblings (Hall and Rossetti 2018; Rossetti and Hall 2015). Having a sibling with a disability in childhood looks very different from the relationship in adulthood. The literature on adult siblings with disabilities focuses primarily on caregiving relationships as it is often siblings who step into caregiving roles for their adult siblings after parents, schools, and other governmental services change or reduce levels of support over time (Sonik, Parish, and Rosenthal 2016; Dew, Balandin, and Llewellyn 2008; Avieli, Band-Winterstein, and Bergman 2019).

Bioecological Theory and Sibling Embedded Systems Framework

The previous three frameworks of sibling relationships focus only on the sibling-sibling relationship in the context of a family. A major theoretical contribution in the field of sibling relationships builds off of Bronfenbrenner's bioecological theory of human development (Bronfenbrenner 1986; Bronfenbrenner and Morris 2006)[49]. The bioecological system theory situates the sibling relationships within the family into a broader system of five hierarchically structures. The (1) microsystems of relationships within the family interact with (2) mesosystems of extended family, school and other support networks. (3) Exosystems include the larger community where the family lives and resources available at a local level while (4) macrosystems are defined as the larger cultural, political, and structural context of governance and policy. The bioecological framework also incorporates elements of the life course framework through the inclusion of the (5) chronosystem (that is, over time). These five systems allow for a more holistic interpretation of sibling relationships that is generalizable to other cultures. In contrast, family systems, family stress, and life course theories are built upon Western representations of family, childhood, and siblings in ways that highlight individual relationships among family members.

[49] Notably in the later years of Bronfenbrenner's scholarship (Rosa and Tudge 2013)

Kovshoff, Cebula, Tsai, and Hastings (2017) modify the bioecological framework to propose the sibling embedded systems framework, which weaves together elements of the established family systems, family stress, and life course theories mentioned above but also emphasizes not only negative outcomes, but the potential (and actual) positive outcomes of having a sibling with a disability.

The bioecological and sibling embedded systems frameworks have rarely been fully utilized in academic scholarship, despite the obvious advantages of situating an individual sibling relationship within the broader cultural and structural systems that may shape it. Typically, research on outcomes is firmly situated at the micro and meso levels of interaction between siblings and within families, prioritizing the family systems, family stress, and individualistic life course theories. Much of the existing research was conducted on middle-class, white American (or European) families, with a few exceptions looking at diverse families in the United States (for example, Harry 2002; McGuire and Shanahan 2006).

None of the theoretical frameworks that I describe have been applied or adapted to study sibling relationships in non-Western countries (Hodapp, Glidden, and Kaiser, 2005), to my knowledge. Research in the fields of sibling relationships and disability commonly neglects to address how theories shape results. As mentioned above, Stoneman argued that sibling relationships has historically been shaped by concepts of an ideal sibling relationship between children without disabilities; this emphasizes different types of sibling relationships *within* a specific (i.e. Western) culture. She also asserts that the Euro-American culture has resulted in idealized sibling relationships that are "warm, supportive, and free of conflict" (Stoneman 2005, p 339). The lack of diverse theory perpetuates western values such as the nuclear family, individualism, and individual rights. A particular kind of childhood is assumed to be the foundation for all research on sibling relationships. The exosystem and macrosystem levels of the bioecological and sibling embedded systems theories are flexible enough to consider different sibling relationships *between* Western and non-Western and/or non-white contexts, but this is hardly ever done in practice.

While my research falls within the family structure and family stress theories presented here, I provide background on family and siblings dynamics, disability, and disability policy in Tanzania (and other Sub-Saharan African countries) to suggest caution towards assuming that the same theories can be adapted to Tanzania without consideration for a holistic policy and cultural context.

3.2.2 Disability in Sub-Saharan Africa and Tanzania

While research on disability and sibling relationships has been developed mostly in European and North American countries, policies that are aimed at the improvement of the lives of persons with disability have been increasing around the world. The United Nations' Convention on the Rights of Persons with Disabilities was drafted in 2006 and has been signed by many African countries, including Tanzania (United Nations 2006). As with many international conventions, the signing of a convention does not translate directly into implementation and enforcement of policies (Aldersey and Turnbull 2011).

One reason for this discrepancy is due to differences in cultural interpretation and definition of disability in Africa. Many of the conventions and Western theories of disability employ a medicalized and diagnostic interpretation of disability (Grech 2015). This medical model often conflicts or imposes a Western standardization of what is means to have a disability or support someone living with a disability on persons living in Tanzania and other Sub-Saharan African countries. Stone-MacDonald and Butera (2012) synthesize from the literature four areas in which cultural beliefs and attitudes about disability in East Africa (which includes Tanzania) parallel or diverge from those in Western or developing countries:

(1) Beliefs about the cause of disability
(2) Attitudes towards persons with disabilities
(3) Treatment of persons with disabilities
(4) Language used to describe disabilities or persons with disabilities

As the literature about interpretation of disability is rich and expansive, I provide only a few examples of how cultural acceptance and rejection of persons with disabilities varies greatly across different cultural groups and the nature of the disability. A person's disability may be considered the result of religious fatalism or divine punishment, though knowledge and belief in medical models of disability and the causes of disability are not uncommon (Nyangweso 2018). Studies also show that individuals with disabilities are not marginalized in East Africa; instead inclusion and integration into broad community and sociality depends the type and severity of the disability the person is living with (Kisanji 1995). For example, a person who is deaf or with difficulty learning often can be included in their community with minimal accommodations the way a person who is blind, with a physical disability, or with a severe difficulty socializing may not be able to. However, inclusion does not mean equal treatment as many children with disabilities are not able to attend schools or are not welcomed by peers (Stone-Macdonald 2012).

Cultural translation, or lack thereof, challenges attempts to create and validate disability statistics and information on persons with disabilities. Small scale ethnographic or qualitative

work best suited for interpreting cultural understanding of disability is not suited for the establishment of cross-cultural statistics on persons with disabilities (Chamie 1989). Moreover, data collection and research that look across countries struggle with variations in definitions of disability (examples include Maulik and Darmstadt 2007; Croft 2013; Fujiura, Park, and Rutkowski-Kmitta 2005). For example, the very language used to describe difficulties and impairments (what are typically called "disabilities" in Western culture) do not always have clear translations of words or concepts (Ogechi and Ruto 2002). The lack of unified statistics on individuals with disabilities and their families make it difficult to capture the interest of policy makers to create and enforce policy.

3.2.3 Families, siblings, and caregiving in Sub-Saharan Africa

As more countries in the developing world create and implement policies to protect and support individuals with disabilities – such as the United Nations Convention of the Rights of Persons with Disabilities (United Nations 2006) – little policy attention is being paid to families of these individuals. Fundamentally, we do not know how many people in developing countries have siblings with disabilities. One way to increase policy attention on individuals with disabilities and their siblings is to establish basic demographic estimates. The lack of basic demographic information and limited cultural adaptations of theoretical interpretations of disability in non-Western contexts are major barriers to conducting research about how outcomes and experiences for siblings of an individual with a disability may differ from people without a sibling with a disability. The lack of research, in turn, makes it harder to advocate for policies and services for supporting the families, and particularly the siblings, of individuals with disabilities.

Policies such as The Convention of the Rights of Persons with Disabilities aim to reduce discrimination against people with disabilities and improve inclusive opportunities. However, many families lack access to services for individuals with disabilities. Research on parents and other caregivers of children with disabilities finds that the stigmatization of disability, the negative outlook by family and community members, and the high cost associated with seeking assistance creates high stress and frustration for caregivers of children with disabilities (Hartley et al. 2005; Bannink, Nalugya, and van Hove 2020; Mbwilo, Smide, and Aarts 2010).

One primary way children with disabilities in developed countries are supported is through inclusive education and school-based support systems. However, children with disabilities in developing countries are less likely to be enrolled in schools (Mizunoya, Mitra, and Yamasaki 2018; Filmer 2008; Bakhshi, Babulal, and Trani 2020). Children who are enrolled in

schools are often not supported with adequately adaptive education including teachers with special education training or other assistive technologies (Mapunda, Omollo, and Bali 2017; Opini and Onditi 2016; Chitiyo and Chitiyo 2007).

Without school-based support networks, parents may struggle to support their children with disabilities. Parents in one study in Tanzania lament that the government did not provide enough support for families facing challenges paying for school fees, and there were few community-based organizations to help children with disabilities and their families (Aldersey 2012). Parents saw the lack of financial support, or inability or lack of knowledge to access funding, in direct contradiction to the government promising to support persons with disabilities. The financial struggles families face to support their children in the absence of governmental financial support puts additional pressure on the whole family, illustrating an aspect of family stress theory.

In the absence of governmental or community-based support, providing care for children with disabilities often falls to mothers (Fotso 2017a) and other children (Chataika and Mckenzie 2013). In Sub-Saharan African countries, the responsibilities of caring for other people in a family (with and without disabilities) often falls to children (Evans 2010). For example, children with parents with HIV take on many caregiving roles for their parents and other siblings (Hunleth 2017). Additionally, in families where a parent has a health problem, children are less likely to be in school due to increased responsibilities at home, overall financial difficulties of the family resulting in an inability to pay for school fees, and occasionally a need for the child to work to support the family (Alam 2015). There are also cultural expectations in many African cultures that older siblings will financially support younger siblings (Trinitapoli, Yeatman, and Fledderjohann 2014).

It is clear that both sibling relationships and cultural interpretations of disabilities in sub-Saharan Africa are very different from the established and prolific literature based on sibling relationships between children (and adults) with and without disabilities in the USA and other developed countries. Research on sibling relationships shaped by a disability within developing countries, such as in sub-Saharan Africa, is limited. A 2017 paper claims to be the first study of the attitudes of adolescent siblings of children with autism spectrum disorder in South Africa (Van der Merwe et al. 2017) and the paper cites two additional articles from South Africa (Opperman and Alant 2003; Hansen, Harty, and Bornman 2016). All three studies rely on small

convenience samples drawn from families where a child with a disability is enrolled in specific "special education" schools located in large cities[50].

One study from Cameroon looks at the impact of having a sibling with a disability at the national level (Fotso 2017b). In this paper, Fotso uses the 2011 Demographic and Health Survey and Multiple Indicator Cluster Survey (conducted jointly) to analyze differences in monthly health care expenditures for children with and without siblings with disabilities. The main finding is that having a sibling with a disability reduces the amount of money spent on a child's health care. This article is an excellent example of how identifying children with siblings with disabilities can lead to important insights into outcomes for children and their families.

The South African examples I describe above are all situated in the family structure theory as they focus on the microsystem relationships and attitudinal changes of a child without a disability towards his or her sibling with a disability[51]. Fotso's study in Cameroon utilizes family stress theory to explore resource distribution among family members. While the studies are located in sub-Saharan Africa, there is little focus on the cultural or political context that could be driving the results. The representation of siblings in these studies is only a small step into exploring the relationship between siblings in sub-Saharan Africa.

3.2.4 Disability and Education Policy in Tanzania

In Tanzania, it is often said that education is the key to life. The educational system in Tanzania consists of pre-primary, six years of primary school, and two cycles of secondary school consisting of four and two years each (United Republic of Tanzania President's Office 2016). Government schools are free to attend, though there is often still a considerable expense to the family to purchase uniforms, books, transportation, other supplies, or boarding accommodations. Most children in Tanzania attend some schooling. In 2016, the net enrollment rate for primary school was 86 percent (UNICEF 2017). However, that figure drops to 33 percent in secondary school with only 3 percent enrolled in the last cycle of secondary. There are vast regional and wealth disparities in school enrollment, though the gap between genders has closed. Policies

[50] Notably, none of the papers specify the race or cultural background of their participants and it is unclear if the families who participated in the study were of a higher socio-economic status or white South African. Two of the papers required the participants to speak English or Afrikaans as the mode of participation.
[51] Van der Merwe et. al (2017) specifically recommends incorporating a chronosystem approach to research about sibling relationships as they find adolescent's attitudes change as the siblings grow from childhood to adolescence.

aimed at educating children do not adequately address social and economic factors that result in children dropping out of school early (Kuépié, Shapiro, and Tenikue 2015).

Like elsewhere in sub-Saharan Africa, persons with disabilities in Tanzania are less likely to be enrolled in schools and those who are enrolled face limited services, lack of inclusive instruction, and few teachers trained with necessary skills (Opini and Onditi 2016; Mapunda, Omollo, and Bali 2017). There are also few specialized schools available for children, and schools that are considered "inclusive" or "integrated" are not always welcoming to students with disabilities (Possi and Milinga 2017).

In addition to not having educational services, families of children with disabilities also lack access to other services such as health care services and/or community-based services to support the child and the family. This can result in high levels of stress for families caring for children at home (Mbwilo, Smide, and Aarts 2010). In a qualitative study in Dar es Salaam, parents shared concerns about their child's wellbeing as they sought to navigate feelings of lack of control and social stigmatization (Aldersey 2012).

As previously mentioned, Tanzania is a signatory to the United Nations Convention on the Rights of Persons with Disabilities. The main disability policy in the country is the National Policy on Disability of 2004 and the Persons with Disability Act of 2010. Both of the policies are aimed at empowering persons with disabilities, expanding the support for individuals and their families, and preventing discrimination in education and employment on the basis of disability. However, there has not much enforcement or accountability of the policies (Aldersey and Turnbull 2011).

In 2009, the government created the National Strategy for Inclusive Education. This plan was renewed in 2017. The purpose of the National Strategy was:

> *… to enhance educational services for children with special needs, by adopting an inclusive approach to policy-making and service delivery planning, ensuring teaching and learning are accessible to all, in particular by developing specific staff capabilities, providing pedagogical support, promoting community participation and ownership of education, and facilitating the use of sign language and Braille (2017)*

The first objective of the National Strategy for Inclusive Education is "Equitable access and participation in at least one year of pre-primary education enhanced for all 5 years old children with particular attention to vulnerable groups." This objective explicitly calls out the need for pre-primary education for all children, particularly vulnerable children or children with disabilities, as a strategy for also increasing the school enrollment of siblings. This particular attention paid to siblings is mentioned once in the entire 2017 policy.

Tanzanian children are often expected to take on care-giving roles and home responsibilities in order to help their families. In another study from Tanzania situated within the family stress framework, Alam (2015) finds children are less likely to be enrolled in school if their father is experiencing a chronic or short-term illness and identifies the main mechanism as likely to be a lack of financial support to pay for school fees, rather than a reallocation of the child's time from school to work or other caregiving activities. The findings from studying educational outcomes for children with a parental illness can be applied further to children with siblings with disabilities[52].

In order to look at the relationship between having a sibling with a disability and enrollment in school in Tanzania, I return to the theories described in the earlier section of this chapter. First, family systems theory considers the relationship between siblings and the roles and responsibilities that a child who is caring for a sibling with a disability may experience differently from a peer without a sibling with a disability. The literature shows that expectations of these responsibilities in Tanzania differ from sibling relationships based on a Western view of siblings. Second, family stress theory suggests that the lack of services for families of persons with disabilities makes it hard for the family to support a child with a disability without significant financial cost. This reduces resources to spend on other children and may affect the child in educational, emotional, or social ways that are frequently situated in family system theory. This theory has been applied to education among siblings in developing countries for other circumstances beyond having a sibling with a disability. Finally, the bioecological lens builds upon these two previously mentioned theories at the microsystem level within the family and chronosystem across the life course. Other elements of the bioecological theory can be seen in the lack of adequate community and school supports for persons with disabilities (mesosystem and exosystem) in Tanzania. I have presented several macrolevel disability and educations policies unique to Tanzania while recognizing the lack of accountability and enforcement of anti-discrimination policies. Finally, I recognize the culturally influences (at the macrosystem) on families in Tanzania shaped by interpretations and definitions of disability. There is not one cultural interpretation across all of Tanzania, but the interactions of children and their siblings with disabilities is influenced by broad and personal experiences within Tanzanian cultures that cannot be captured in this analysis.

[52] Research in other developing countries finds that resource allocation among siblings may be uneven based on gender, family size, or availability of government support for school fees (Barrera-Osorio, Linden, and Perez-Calle 2008; Banerjee and Duflo 2011).

To better understand sibling relationships in Tanzania, the first step of the analysis is to identify children who have disabilities and who have siblings with disabilities and quantitatively describe their prevalence in the population. Second, my analysis on school enrollment in Tanzania is situated within the family systems and family stress theories. This analysis aims to take a first step towards reflecting on how theories of sibling relationships can be applied or adapted to Tanzania. In order to properly incorporate the cultural, policy, and historical context of Tanzania, I argue a bioecological and sibling embedded framework should be used to better understand the intersection of disability, childhood, and sibling relationships. This, however, is beyond the scope of this chapter.

In order to apply either theoretical framework to Tanzania, it is important to consider how different representations of disability and childhood and cultural factors shape sibling relationships in order to be open to finding positive and negative outcomes. One contribution of this chapter is to demonstrate a method of identifying children with siblings with disabilities in order to draw policy attention to needed support for persons with disabilities and their families.

3.3 Data

Tanzania 2012 IPUMSI-International Census

Census data is ideal for constructing population estimates of the total number of children who have a sibling with a disability as this is a relatively rare occurrence that may be difficult to adequately capture in nationally representative samples. The sample of the Tanzania 2012 census data was obtained from IPUMS-International (IPUMSI) (Minnesota Population Center 2019 & National Bureau of Statistics). The IPUMSI data is a 10 percent sample of private households drawn using systematic sampling from the Tanzanian 2012 census. All estimates and analysis use the differential weights provided by Tanzania's National Statistics Office to IPUMSI.

IPUMSI census data is microdata. Microdata provides information about each individual person separately. Additionally, the IPUMSI census data organizes individuals into households. This is ideal for studying family structures and family relationships. The primary variable used to identify family structure identifies each individual's relationship to the "head of the household"[53].

[53] The conceptualization of a "head of household" is critically debated in demography and population studies. In the Tanzanian 2012 census, the head of household is defined as "a person responsible for a particular household or a person recognized as the leader by other household members in a particular household" (IPUMSI). Typically, this is automatically defaulted to the eldest member, the highest earning

This often allows for children to be identified with their parents from their stated relationship to the head of household and for sibling relationships to be identified among those children of the household head who are reported in the same household. However, there are many types of family relationships that are not captured by census data, such as multi-generational households (discussed below).

Disability in the 2012 Tanzanian Census

Census data has a long and complicated history of asking questions about disability status. Initially, disability status in census data was capture by a simple binary question, "Does [this person] have a disability? Yes or no". Occasionally, the type of disability was specified in broad categories such as "blind", "deaf", or "mental disability".

In 2002, the United Nations Washington Group (WG) reviewed how censuses and surveys worded questions about disability in response to concerns by disability advocacy and international organizations. Disability studies and policy aimed to expand the definitions and conceptualization of disability status in censuses and surveys to be more inclusive of social barriers and limitations instead of only impairments. Recommendations out of the WG include a series of questions that focus on aspects of life and varying levels of difficulties (Washington Group on Disability Statistics 2020). For example, instead of asking "Is [the respondent] blind? Yes or no?" a WG question asks "Does [the respondent] have difficulty seeing, even when wearing glasses?". The recommended answer categories allow for a 4-point scale of "No difficulty", "some difficulty", "a lot of difficulty", and "cannot see at all". The variation in these answers allows for more nuance. Being blind or answering "cannot see at all" may lead to certain requirements and needs of the person in order to live an inclusive life, but previous censuses may have underreported the number of persons who experience some or a lot of difficulty in day-to-day seeing. Several countries incorporated these recommendations into their census enumerations since the early 2000s.

The Tanzania 2012 census utilizes the latest recommended questions posed by the WF to measure disability[54]. Five questions ask about difficulty seeing even while wearing glasses, difficulty hearing even while wearing a hearing aid, difficulty walking or climbing stairs,

individual, or the husband or father figure in the household, though there is no way to verify exactly how an individual household selected a "head" to answer the census questions.

[54] An excellent resource for more information on the Washington Group and comparable measurement of disability variables found in (Altman and Barnartt 2006).

difficulty remembering or concentrating, and/or difficulty in self-care, including washing and dressing.

To identify persons with disabilities, I use a binary variable created by IPUMSI to identify persons who report "a lot of difficulty" or the "inability to do [X] at all" for any of the five above mentioned areas of the WG, and one additional "other disability" variable provided in the census data. This binary variable reduces the nuance provided by the WG conceptualization of disability. However, WG documentation and reports suggest establishing a "cut-off point" to identify individuals with severe disabilities in order to make the WG recommended questions comparable to other disability statistics, in order to better communicate to policy makers the sub-population requiring more direct and intense services and policies (Washington Group 2018, page 51). Future research will consider alternative definitions of disability status.

School enrollment in the Tanzanian Census

The primary variable of interest to measure school enrollment is a question that asks about persons over the age of four, "Is [the respondent] currently attending, partially attended, completed or never attended school?" Using this variable, I identify children who are currently attending school. It is not possible using census data to assess the frequency of school attendance, the quality of education, or the reason for not being enrolled, all of which may influence the results of the analysis (Bold et al. 2017; Levison, DeGraff, and Dungumaro 2017).

The IPUMSI sample of Tanzanian 2012 census data includes only private households. This is a limitation to the analysis of school enrollment as many secondary schools and some private elementary schools in Tanzania are boarding schools. The 2012 Census Day was August 26; thus, many school children would have been in the middle of an academic semester[55]. Additionally, the Tanzania census was a de Facto census. This means that children attending boarding schools, who are not living with their family on the night of the census, would therefore not be counted as being part of the household. It is very difficult to find statistics on how many children attend boarding schools. Boarding schools and other types of institutional group living quarters are not available in the IPUMSI census data for Tanzania 2012. The implications of the lack of institutional data for persons with disabilities will be discussed below.

[55] School breaks in Tanzania are typically in June and December

3.4 Method

In this section, I return to my two research questions to be addressed in this chapter. First, I will establish demographic estimates of the number of children in Tanzania living with a sibling with a disability. Second, I will compare school enrollment status of the children identified as having a sibling with a disability to children who do not have siblings with disabilities.

3.4.1 Population estimates of children with siblings with disabilities

The primary objective of this chapter is to establish estimates of the number of children in Tanzania living with a sibling who has a disability. In this section, I describe the process of identifying siblings within a household and some of the limitations in the 2012 Tanzanian census data. Then I discuss several operationalizations of sibling relationships of children with and without disabilities.

Identification of the Child

In this section, and through the chapter, the unit of observation is the child. Each child within a household is individually assigned a sibling status based on their relationship with other children. I set several inclusion criteria to establish accurate representation of sibling status.

First, as defined by the United Nations Conventions of the Rights of the Child, I consider children to be people under the age of eighteen (United Nations 1990). I recognize that the analysis in this chapter can be applied to adult siblings as well as child siblings. However, children under the age of 18 are more likely to be living in the same home as their parents and siblings than are adults[56].

Second, I classify a child as being a sibling only if he or she is living at home with parents. This means that the individual must be reported as being the "son or daughter" of the head of the household. The restriction of children living with parents also allows me to identify sibling sets. In the Tanzania 2012 census instructions provided by IPUMSI, "sons and daughters" refers only to biological children of the head of the household. This also excludes individuals who

[56] I recognize that some children under the age of 18 may be living with a sibling with a disability who is over 18 and living at home. The age limitation allows this paper to be situated with the existing literature of childhood sibling relationships. As a sensitivity analysis, I conducted the school enrollment analysis that follows with all children of the head of household (regardless of age). The results were substantively similar to results when restricting analysis to only children under the age of eighteen (18). Additional research needs to be conducted with adult siblings and adolescent siblings, but that is beyond the scope of this paper.

are the child of the spouse of the head but not of the head of household (aka "step-children")[57]. It is not possible to know to what extent this is accurately reported by households. Limiting the scope to children living with their parents also allows for a more accurate identification of siblings (biological or otherwise). It must be recognized that many children in developing countries live in multi-generational, blended, and other non-nuclear families. Many African children live in households with siblings, half-siblings, cousins, foster siblings, and other non-related children. The presence of a child with a disability in a household may affect not only his or her biological siblings, but other children living in the household (Fotso 2017b).

It is possible, and likely, that many children around the world are not living with their parents for a number of reasons (poverty, educational opportunities, personality conflicts). Any children not living with their parents (for example, living with grandparents or aunts and uncles) are not included in this analysis. Though not specified in the census instructions provided by IPUMSI, it is also possible that the head of household is determined to be the oldest member of a household. This might mean that a grandparent is the head of household and any grandchildren living in the household, with or without their parents, would not be captured in this analysis. It is regrettable to lose these rich and complex family structures in this categorization. The number of children found in this analysis will be a conservative estimate of the total children who may be interacting in some way with a sibling or other close relative with a disability.

My third inclusion criterion is that the child must be living in private household. As described above, some children may be living in group accommodations such as boarding schools or other institutions. Such children are not captured in this analysis because these data are not available in the IPUMSI census sample. Similar to children living with other relatives, it is unknown how many children have siblings who are living in group quarters.

The lack of data on group quarters likely leads to undercounts of persons with disabilities. Historically, persons with disabilities were institutionalized in many western and non-western countries (Fakhoury and Priebe 2002). In Tanzania, most people with disabilities were cared for by families in the home (Whyte 1991). Today, some children with disabilities who do

[57] As the spouse of the head of household is more often the wife, it may be that her children are excluded from this analysis if the household roster strictly follows the inclusion criteria of only "biological sons or daughters" of the head. The degree in which differential treatment of step-children influences the results in the analysis depends on the relationship between non-biological children and school enrollment. One example suggests that step-fathers in the hunter-gather Hadza tribe in Tanzania spend less time with step-children than they do with biological children (Marlowe 1999). More recent research from another area in Tanzania suggests there are small but not significant differences in educational outcomes of fostered children and biological children (Hedges et al. 2019).

attend school – only a small fraction of the total children with disabilities—attend specialized schools (Sida 2014). It is not fully known how many children with disabilities attend boarding schools, though there are examples of specialized day schools (Stone-Macdonald 2012). The impacts of having an institutionalized sibling on a non-disabled child are not well studied in any context. This analysis cannot identify children who are living at home but have a sibling with a disability living elsewhere—either in group quarters or with another private household.

Identification of Siblings with Disabilities

Based on the above inclusion criteria, I identify a sample of children of the head of the household under the age of 18. If there are multiple children in a home who meet these criteria, they are considered siblings. As the unit of analysis is the individual child, the number of siblings a child has is one less than the total number of included children in the household; the child of reference is not counted among the siblings. For example, in a family with three children who meet the inclusion criteria, each child has two siblings. If only one child is living at home with their parents (i.e. there are no other children categorized as siblings reported in the household), it is possible that other biological siblings are older than age 18 or are living elsewhere. These "only children" are not included in the analysis, though they are identified in the descriptive results.

The main descriptive analysis estimates the number of children who have at least one sibling with a disability. This binary dichotomization of sibling status is complicated when taking into consideration each individual child's disability status. As much of the literature on sibling relationships focuses on children who do *not* have a disability, I present the descriptive statistics on sibling status separately for children who have a disability (in accordance with the binary disability measure described above) and children who do not. Each child in the data also has an individual disability status, as described above. This allows me to correctly identify children who have and do *not* have disabilities as independent from their sibling status. The combination of disability status and sibling status results in four major categorizations, as presented in Table 3.1.

Every child in the Tanzanian 2012 census data who is under the age of 18 and living with a parent is assigned one of these categories. All children in families where none of the children have disabilities, (Category A) are assigned to the same categorization. Category A constitutes the most common experience for children. Children in category B (child does not have a disability and at least one sibling has a disability) are the focus of most of the literature on sibling relationships between children without disabilities who have a sibling with a disability.

Table 3.1: Categorization of disability and sibling status for reference child

Category	Disability Status (Does the child have a disability?)	Sibling Status (Does the child have at least 1 sibling with a disability?)
A	No	No
B	No	Yes
C	Yes	No
D	Yes	Yes

The identification of children with disabilities allows for nuanced categorization of children with different experiences within families. Category C (child has a disability, but none of his or her siblings have disabilities) identifies children who are the only child in the family with a disability. The siblings of these category C children are categorized as B as they do not have disabilities, but their sibling does. Category D children (children who have disabilities and have at least one sibling with a disability) are found in families where multiple children have disabilities; children who are siblings with a child in category D but who do not have disabilities themselves would be in the B category. To date, there is limited research on the experiences of the sibling with the disability (Begum and Blacher 2011); the experiences of children in category C or D are often ignored despite children in category D also being siblings to a child with disabilities themselves. I present the descriptive statistics for each of these categories separately to contribute demographic statistics on children with disabilities and also siblings of children with disabilities in Tanzania. Though in the analysis on school enrollment, I will present results that compare children in category B to category A as is most common in the literature[58].

[58] As discussed previously, this chapter is an analysis on sibling status and school enrollment, not school enrollment for children with disabilities. The number of children who have a disability who also have a sibling with a disability is a very small fraction of the total children in Tanzania. In order to avoid confounding the results, I have opted to omit children with disabilities from the analysis on school enrollment. That is, the analysis on school enrollment only compares children without disabilities who *do not* have siblings with disabilities (Category A) with children without disabilities who *do* have siblings with disabilities (Category B). I conducted sensitivity analysis and found when including children *with and without* disabilities in the analysis on school enrollment (Categories A, B, C, and D), there was no significant difference in school enrollment for children with siblings with disabilities (Categories B and D) compared to children who have no siblings with disabilities (Categories A and C).

Additional Operationalization of Sibling and Disability Status

The operationalization of siblings into a binary sibling status (does the reference child have a sibling with a disability or not) mirrors much of the way the existing literature conceptualizes the experience of having a sibling with a disability. The presence of one person with a disability in a household is enough to completely change the experience of all household members, particularly siblings, compared to households where no children have disabilities. In addition to this binary operationalization, I construct alternative measures of sibling status to explore the quantity and compositional effects of sibling relationships within a household.

As much of the existing literature on sibling relationships focuses on sibling pairs or dyads (described above), there is little known about how having multiple siblings in a family affects children. If a child without a disability has multiple siblings with disabilities, that experience may be different from the experiences of a child with only one sibling with a disability. Considering family stress theory, a family with multiple children with disabilities may have fewer resources to allocate among all children, depending on the needs of the children with disabilities.

Another potential factor affecting children's experiences may be the composition of siblings with and without disabilities (Burke, Arnold, and Owen 2015). A child who only has siblings with disabilities may experience sibling relationships differently from a child who has both siblings with disabilities and siblings without disabilities. Children with only siblings who have disabilities do not have a counterfactual experience of what a sibling relationship with siblings without disabilities is like.

In order to explore these additional considerations about the quantity or composition of sibling with and without disabilities within a household, I construct five variables for each reference child as presented in Table 3.2. As described above, the *binary* variable identifies children who have at least one sibling with a disability. Next are two variables to identify the number of siblings with disabilities a child has. The first is a *count* variable presented as a continuous numeric variable that counts the number of siblings with disabilities. This count variable can be displayed in analysis either as a continuous variable or as a categorical variable that allows for each number of siblings with disabilities to be represented in the analysis. In this 2012 Tanzania sample, children have between 0 and 7 siblings with disabilities, as will be described below. The second type of quantity variable is an expansion of the binary variable that identifies if the reference child has zero, one, or multiple siblings with disabilities. The primary way this *categorical quantity* variable is operationalized is by looking at children with 0, 1, or 2

or more siblings with disabilities, but I also consider children with 0, 1, 2, or 3 or more siblings with disabilities in select analyses.

The second group of variables captures the nuance of the composition of siblings (of the reference child) with and without disabilities. The *categorical composition* variable identifies if the reference child has no siblings with disabilities, some siblings with and some without disabilities, or all siblings with disabilities. This is different from the *categorical quantity* variable because the total number of siblings is not considered. For example, a child with a single sibling who has a disability would have one sibling with a disability and at the same time, all of their siblings have disabilities. Each variable considers a slightly different question about the experience of the sibling; the *categorical quantity* variable includes experiences of multiple siblings with disabilities while the *categorical composition* variable identifies the experience of having both siblings with and without disabilities. Finally, in order to quantify the composition of the siblings, I create a *proportion* variable that identifies the proportion of siblings with disabilities out of the total number of siblings.

Each child within a household is individually assigned a value for each of these variables. Example households are shown in Table 3.3. Household 1 shows two children and neither child has a disability. This type of household is considered the majority experience as none of the siblings have disabilities. Household 2 is an example of a household where one of the three children has a disability. The two children in household 2 without disabilities share characteristics as they both have one sibling with a disability and one without. Notice, the child with the disability does not have any siblings with disabilities. Household 3 shows four children where two have disabilities and two do not have disabilities. These children all have at least one sibling with a disability but they also all have the experience of sibling relationships with siblings with and without disabilities. Finally, household 4 shows two children where one has a disability. For the child without a disability, “all” of their siblings have a disability.

3.4.2 Logistic Regression Model

As discussed above, my second research question explores the impact of having a sibling with a disability on children without a disability, specifically comparing school enrollment for children with and without siblings with disabilities. The focus of this analysis is school enrollment of children who are siblings of children with disabilities and not on the school enrollment of children with disabilities. Therefore, I limit the analysis sample to children who

themselves *do not have* a disability but are living with siblings who may or may not have disabilities (Category B compared to Category A from Table 3.1). I use multivariate logistic regression to compare school enrollment for children who have siblings with disabilities to children living in families where none of the children has disabilities. In accordance with the inclusion criteria presented above, I identify all of the children of the household head under the age of 18 in order to identify households where there are children without disabilities who are living with siblings with disabilities. I eliminate "only children" who do not have other siblings living in the same household. I limit this analysis further to reference children who are at least age seven as this is the age where children are typically enrolled in school in Tanzania; the siblings with disabilities can be any age under 18.

I present results from the Tanzania 2012 census using multivariate logistic regression to find the probability that a child, as described by the parameters above, is currently enrolled in school. School enrollment is measured as being currently enrolled in school at the time of the census. Children who were never enrolled in school, or who attended school in the past but were not enrolled at the time of the census, are considered not in school.

A series of logistic regressions for school enrollment are estimated of the general form:

$$\ln\left[\frac{p}{1-p}\right] = \alpha + \beta_1 * X_1 + \beta_2 * X_2 + \cdots + \beta_k * X_k$$

where *p* is the probability that a child is currently enrolled in school. The non-linear logit transformation is estimated as the linear function of an intercept (α) and set of independent variables (*X*) including sibling relationships, demographic characteristics, and household characteristics and their corresponding parameter estimates (β) (Menard 2002). Results are presented in odds ratios[59].

The main independent variable of interest is the effect of having a sibling with a disability. I operationalize this model to test the five conceptualizations of sibling status presented above [Table 3.2]: (1) a *binary* measure of having a sibling with a disability, (2) the number (or *count*) of siblings with a disability, (3) having none, one, or multiple siblings with disabilities (*categorical quantity*), (4) the *categorical composition* variable describing the composition of the child's sibling as none, some, or all having disabilities, and (5) the *proportion* of siblings with

[59] The ease of interpretation of odds ratios is one reason to use logistic regression in this analysis (Hosmer, Lemeshow, and Sturdivant 2013). Odds ratios allow for the comparison of results across the different operationalizations of sibling status in a non-linear, non-symmetric distribution. I conducted the analysis using probit models and found that while the coefficients and standard errors for almost all results were slightly smaller, the direction, significance, and magnitude of the results were the same.

disabilities. Each of these independent variables is tested in separate models. In each of the models, each child – under the age of 18, living with family, and without a disability themselves – is the unit of observation. Children who do not have any siblings with disabilities are the comparison group.

For the other independent variables (*X*), I introduce a series of control variables of demographic characteristics into the models: child's age (measured continuously between 7 and 17), child's sex, and number of total siblings living in the household (measured continuously)[60]. Additional independent variables (*X*) consider socio-economic characteristics of the household through urban or rural status and wealth quintiles created from an asset index of household characteristics and utilities: electricity, radio, TV, metal roof, finished walls, improved cooking fuel, cell phone, flush toilet, access to piped water, refrigerator, car, and owns the home[61]. I created wealth quintiles using a principal component analysis of household assets present in the census (Loveton 2011; Filmer and Scott 2012; Vyas 2006). Standard errors are clustered at the household level to account for characteristics shared by children living in the same household[62]. This analysis utilizes the person weights provided to IPUMSI by the Tanzania National Bureau of Statistics.

3.5 Results

3.5.1 Descriptive Statistics of 2012 Tanzanian Census

In the 2012 Tanzanian census, 50 percent of the population is under the age of 18. Of the children under the age of eighteen, 71.5 percent are reported to be the child of the head of household. The other 23.4 percent are reported as "other related" members of households; thus, these children are living in households headed by someone who is not a parent. Of these "other relatives", 14.6 percent are reported as grandchildren while the remaining "other related" children may be cousins or nieces/nephews, though this level of detailed family relationship is not

[60] I also ran all models using a control variable for total size of household, but this variable was highly correlated with number of children of the head of household in the household and produces essentially identical results. I have opted to use the number of children in the household instead of the total household size.
[61] This follows the wealth index used by Demographic and Health Surveys (citations mentioned in text).
[62] Clustering standard errors at the household level increases the potential for correlated error among siblings without disabilities living in the same household. I address this issue in a sensitivity analysis using the household as the unit of analysis instead of each child, described below.

available in the census. Unrelated children (4.5 percent of 0-17-year-olds) might be foster children, live in domestic helpers, or children of unrelated families living together. A small number of children (0.69 percent of the total 0-17-year-olds) report being either the head of a household or the head's spouse. As shown in the population pyramid in Figure 3.1, the vast majority (83 percent) of persons being reported as the child of the head of household were under the age of eighteen.

Children with disability

Of the 1.6 million children under the age of 18 who are reported as the biological son or daughter of the head of household in the IPUMSI sample of the 2012 Tanzanian census, approximately 1.07 percent (n=17,488; weighted n= 170,0878) have a disability characterized as having "a lot of difficulty" or "not able to do X at all" in at least one of the six disability categories provided in the census, as shown in Table 3.4. This estimate is lower than the commonly cited figure that 5 percent of children globally have a disability (Thompson 2017). Table 3.4 shows the five WG categories – seeing, hearing, walking, remembering, and self-care – and the four levels of difficulty reported for children below 18 years old. The vertical axis shows only the proportion of children characterized in the top 1-2 percent. For all of the different areas of difficulty, hearing has the most children reported as having some difficulty. More children are reported as being unable to walk and remember compared to seeing or hearing. The value of using the WG difficulty ratings allows for nuance in the level of a perceived difficulty or disability. However, only children with "a lot of difficulty" or "unable to do X at all" are included in this analysis[63]. Compared to the other variables, there are more individual children who are reported to not be able to care for themselves at all (self-care). This is due to a slight difference in the universe; the census question was not asked to children ages 0 to 3 years. This was done to avoid conflating the care needs of young children with needs of an individual with disabilities. For example, an infant would have difficulty feeding, dressing, and cleaning themselves without it being considered a disability.

As discussed previously, disability statistics, even in census data, are difficult to validate as being accurate. The numbers of children with a disability in Tanzania reported in this chapter are likely is an underestimate. In the census, 62.6 percent of all households include a person under the age of 18 who is a child of the head of household and 1.6 percent of all households

[63] As mentioned before, my future research will look at the WG nuanced disability measures beyond the binary measure constructed by IPUMSI.

have a child with a disability. Only 0.04 percent of households have multiple children with disabilities living in the same household.

3.5.2 Population estimates of children living with siblings with disabilities in Tanzania

Using the 2012 census, I identify the number of children who have disabilities and the number of children who are the siblings of a child with disabilities. The primary focus of this chapter is to understand how many children, who may or may not themselves have a disability, have at least one sibling with a disability.

Children with siblings

Table 3.5 shows the number of children age 0-17 who are living at home and reported as the child of the head of the household, by disability status. The results are presented using four of the operationalizations of sibling status: as a binary variable, a count variable, a categorical quantity, and a categorical composition. The population estimate is calculated using the weights provided by IPUMSI to scale the sample to the population level for Tanzanian in 2012.

In the Tanzanian 2012 census, children are living with an average of 2.77 siblings (3.77 average children of the household head in the household)[64]. Ten percent of children were reported as living without siblings under the age of 18 ("only children"). Of the children in my sample *who have siblings*, 2.7 percent of these children have a sibling with a disability (n=39,506; weighted n= 378,878). While most of these children do not themselves have disabilities, an additional 11,126 children have disabilities, but none of their siblings have disabilities. In total, 3.46 percent (n = 50,632, 3.07 percent, weighted n = 487,000) of the population of children 0-17 reported as the son or daughter of the head of household either have a disability or have a sibling with a disability. When combining the number of children with disabilities and the number of children who have siblings with disabilities, this number – almost half a million – gives us a broader estimate of the size of the population of interest compared to statistics at the household level.

[64] This is lower than the overall fertility (number of children born to a woman) – 5.4 children per woman (National Bureau of Statistics (NBS) Tanzania and ICF Macro 2011)— because of childhood deaths, children living outside of the home, and children over the age of 18 at the given time of the census.

Quantity and composition of siblings with disabilities

In order to further explore the results of the various operationalizations of sibling status presented in Table 3.5, I describe in further detail results comparing children with siblings (i.e. excluding those who are only children).

The *count* variable (as described in Table 3.2) shows the number of children who are living with at least one sibling with a disability; it describes how many siblings of the reference child have disabilities. Most (85.5 percent) of the children living with siblings with disabilities only have one sibling with disabilities, as shown in Figure 3.2. The number of children living with multiple siblings with disabilities decreases as the number of siblings with disabilities increases.

Figure 3.2 shows that only 10 percent have multiple siblings with disabilities. Children living with disabilities are more likely to be living with multiple siblings with disabilities than children without disabilities. This supports other research that disabilities tend to cluster in families (Shandra et al. 2012). Among children with disabilities who have at least one sibling with a disability, 23 percent have multiple siblings with disabilities.

Figure 3.3 shows the composition of siblings for children with and without disabilities, representing the *categorical composition* variable. These figures include the proportion of children who do not have siblings and do not have siblings with disabilities for comparison. The results show that children with disabilities are more likely to have siblings with disabilities. Most children without disabilities also do not have a sibling with a disability (the majority experience and main comparison group in this analysis). The composition of families also shows that most children who have a sibling with a disability also have siblings without disabilities (the category "some").

In the *categorical composition* variable, all children with *at least one* sibling with a disability would be categorized as having some or all siblings with disabilities. Using information from the *count* variable, additional analysis (not shown) clarifies the vast majority (81 percent) of the children who have siblings who all have disabilities are reporting only one other sibling in the household (therefore 2 children in the household, one with a disability and one without). But more children with all siblings with disabilities have a larger number of siblings with disabilities than children who have some siblings with disabilities and some siblings without disabilities. This again shows the clustering of disability in families, but also is a mathematical property of identifying children's disability status and sibling status independently. For example, if there are three children in a family and all of them have disabilities, they would all three appear in the data as having a disability and also having 2 (or all) of their siblings with disabilities. In contrast, a

family where one child has two siblings with disabilities would report having "all siblings with disabilities" but the two children with disabilities would each report as having "some siblings with and some without disabilities" (such as in Household 3 in Table 3.3).

Finally, Figure 3.4 shows that the composition of siblings with and without disabilities as a histogram of the *proportion* variable for all children with at least one sibling with a disability (notice, zero is empty as none of these children would have zero siblings with disabilities). Most children have 25 percent, 33 percent, or 50 percent of their siblings having disabilities, results similar to the composition of siblings in Figure 3.3: most children who have some siblings with and without disabilities have one sibling with a disability. For example, a household with four children where one of them has a disability means that the children without disabilities in the family would have one of their three siblings (1/3) reported as having a disability.

These descriptive findings show children with siblings with disabilities vary in the quantity and composition of their sibling relationships. Some children are living with one sibling with a disability, a situation comparable to research conducted in the Western context as a sibling pair where one child has a sibling with disabilities and the other does not. Less commonly, other children in Tanzania are living in large families with many children with and without disabilities. Children with disabilities are also living with other children with disabilities; this is an important theme to explore that is beyond the scope of this paper.

3.5.3 Logistic Regression

School enrollment

The second research question is whether having a sibling with a disability impacts school enrollment for children. This analysis is restricted to children between 7 and 17 years of age to account for the ages of typical enrollment in primary and secondary school in Tanzania. Among children 7 to 17 years of age, 78 percent were enrolled in school at the time of the census. Table 3.6 shows descriptive statistics for the demographic and household control variables for children currently enrolled in school compared to children not enrolled. Children enrolled in school tend to be slightly younger, reflecting the higher primary school rates of enrollment compared to secondary school. Children from wealthier households make up a larger proportion of the enrolled students. Among enrolled students, there are slightly more boys enrolled in school than girls. The descriptive statistics for children with disabilities look fairly similar to children with disabilities.

Figure 3.5 shows three of the different operationalized variables (*binary*, *count*, and *categorical composition*). Of children who have at least one sibling with a disability (*binary*), the overall percentage of school enrolled children with a sibling with a disability is very similar to overall proportion of children enrolled in school who do not have a sibling with a disability. Only for children who have multiple siblings with disabilities (5, 6, and 7 siblings) do the results differ greatly and these results are driven by comparatively lower sample sizes as shown in the results for the *count* variable. Based solely on these descriptive results, it does not appear that school enrollment for children with a sibling with a disability is greatly different from school enrollment for children in families where no children have disabilities, when not controlling for any additional variables.

School enrollment for children 7 to 17 with disabilities is 54 percent. As expected, this proportion is much lower than the enrollment rate for children without disabilities (78 percent). In the analysis conducted below, I do not include children who have disabilities; my focus is on children without disabilities who have a sibling with or without disabilities[65].

Logistic Regression Results: Binary outcomes with and without controls

I calculate the odds ratios for being enrolled in school for children between ages 7-17 who do not have a disability. I exclude children who do not have any siblings living at home ("only children") from the analysis.

Table 3.7 shows three models comparing children who have a sibling with any disability to children who do not have any siblings with disabilities using the *binary* variable. Model I includes no controls, Model II includes demographic characteristics of the child, and Model III includes the household characteristics. In Model I, having a sibling with a disability significantly decreases the odds (OR= 0.81) of a child without a disability being enrolled in school. This is similar to the descriptive findings in Figure 3.5. Models II and III show that this finding loses significance once demographic and household characteristics are included in the model. The demographic control variables show similar patterns to the descriptive statistics: older children and children from larger families are less likely to be enrolled in school and females are slightly more likely to be enrolled in school. Finally, children living in urban areas and from wealthier households are more likely to be enrolled in school than rural children and poorer children. These

[65] I did however conduct a regression analysis of children with disabilities age 0-17 and found that the odds that children with disabilities are enrolled in school is 67 percent lower than children without disabilities, as presented in the appendix.

results suggest that children with siblings with disabilities are less likely to be enrolled in school, but these effects are moderated by other characteristics also associated with school enrollment.

Other operationalization of sibling status

In order to explore quantity and composition effects of having multiple siblings with or without disabilities, I conduct the analysis of school enrollment for the other operationalizations of sibling status. Table 3.8 shows that odds of being enrolled in school for seven different models of operationalization without control variables. All models use children who have no siblings with disabilities as the reference group. Model I shows the binary results as presented above. Model II and Model III show modifications on the binary variable to include categories for two or more siblings with disabilities and 2 or 3 or more siblings with disabilities, respectively. Model IV operationalizes the number of siblings with disabilities a child has as a continuous numeric variable and Model V uses the same variable but as an ordinal variable. Model VI represents the operationalization that highlights the composition of the family considering children who have some siblings with disabilities and some without disabilities compared to having all siblings with disabilities. And the final model (VII) looks at how the likelihood of school enrollment is affected by the proportion of siblings with disabilities each child has.

In all of these operationalizations, having one sibling with a disability decreases the reference child's odds of being enrolled in school by approximately 20 percent. As shown in Models II, III, IV, and V, the more siblings with disabilities a child has, the less likely they are to be enrolled in school. When considering the composition of siblings, children with some siblings with disabilities are less likely to be enrolled in school compared to children with no siblings with disabilities, but there is no significant difference for children with all siblings with disabilities.

However, Table 3.9 shows that once the demographic and household characteristics are included in the models, the significant decrease in likelihood of school enrollment is eliminated. In most of the models, having a sibling with a disability does not appear to affect school enrollment of children who do not have siblings with disabilities; and if there is negative directional affect, it is not significant[66]. In Models II, III, and V, having multiple siblings with disabilities does decrease the probability of school enrollment compared to a child without any siblings with disabilities, but this is not a significant difference. The only Model that has any significant results is Model VI for the children who have only siblings with disabilities (all of their siblings have a disability) (odds ratio = 0.75).

[66] Coefficients for the control variables for each of the models are virtually identical to the control variables found in Table 3.

When considering the interpretation of the results from Model VI, I return to my argument for why it was important to consider the composition of siblings within a household. The *categorical composition* variable focuses our attention on children who only have siblings with disabilities, compared to children who have at least one sibling who does *not* have a disability. Children who have siblings who *all* have disabilities make up 0.2 percent of the population of children without disabilities [Table 3.5]. Most children with all siblings with disabilities either have one or two siblings with disabilities. When comparing across of the Models in Table 3.9, the significance of Model VI for only children who only have siblings with disabilities may be explained using Family Systems and/or Family Stress theories. If there are multiple children without disabilities in a household ("some"), it is likely the work of caring for a sibling with a disability or other household responsibilities are distributed across multiple children. Therefore, the reference child who is the only child at home without a disability may be less likely to be enrolled in school due to household responsibilities that fall entirely to them.

Despite the significance of Model VI, consistently insignificant relationship between school enrollment and sibling status across various operationalizations of sibling status is striking. It suggests that the likelihood of being enrolled in school is not dependent on having a sibling with a disability. This null finding is important because it suggests that having a sibling with a disability in Tanzania is not a major contributor to school enrollment. Additionally, the results presented in Model VI present evidence that the composition of siblings a child has may be more important than the number of siblings with disabilities. Finally, the lack of negative significant results in Table 3.9 suggests that perhaps the literature on negative outcomes for children with siblings with disabilities that is common in Western literature is not applicable to Tanzania.

3.5.4 Additional Sensitivity Checks of Results

The main results shown in Table 3.8 and Table 3.9 show that the effect of sibling status on school enrollment is no longer significant when demographic and household control variables are included in the models for almost all operationalizations of sibling status. In order to further explore the validity of these results, I conduct a series of sensitivity analyses. First, I conduct a regression analysis of the presence of children with disabilities at the household level. Analysis at the household level illuminates household level dynamics that may be influencing why some children are enrolled in school and not others. Second, I present a series of results that return to

the child as the unit of analysis to explore the control variables of the wealth quintiles and gender of the reference child and the sensitivity of the dependent variable of school enrollment.

Household Level Analysis

All analyses thus far in the chapter present results with the child as the unit of analysis. This allows for children within a house to have different sibling statuses, disability statuses, and school enrollment. However, children clustered within families may share similar observed and unobserved characteristics. For children without disabilities who have siblings who all have disabilities, this is not a problem as the child without the disability will be the only child from the family represented in the analysis. However, for children who have no siblings with disabilities or a mix of siblings with and without disabilities, each child without a disability in the family will be counted.

Many of the theories of sibling relationships consider the sibling relationship to exist within a family structure. Siblings may influence each other on an individual level in family systems theory, but family stress theory suggests that having a child in the family with a disability may impact decisions made at the family level. Families with one child with a disability may make different decisions about school enrollment choices for all children in the family compared to families of similar size, wealth status, and gender composition without a child or children with disabilities.

I test the association between school enrollment and having a sibling with a disability at a household level, using the following regression model:

$$y = \alpha + \beta_1 * X_1 + \beta_2 * X_2 + \cdots + \beta_k * X_k$$

where *y* is the proportion of children (converted into a percentage for interpretability) enrolled in school. The main independent variable of interest (X_1) is an indicator as to whether a child under the age of 18 with a disability is living in the household. While the sibling with a disability could be any age, I limit the analysis to only identify school eligible children (age 7-17) to calculate the proportion of children enrolled in school out of the total eligible children. In Table 3.10, Model 1 shows the results for *y* where the proportion of eligible children (age 7-17) includes both children with and without disabilities.

$$y = \frac{all\ enrolled\ children\ (7-17)\ in\ household}{all\ children\ (7-17) in\ household} * 100$$

In Model 2, I include only children age 7 to 17 who do not have a disability in the numerator and denominator.

$$y = \frac{all\ enrolled\ children\ (7-17)\ \boldsymbol{without\ disabilities}\ in\ household}{all\ children\ (7-17)\ \boldsymbol{without\ disabilities}\ in\ household} * 100$$

This allows me to isolate the association of having a sibling with a disability on children without disabilities. As in the main analysis, the definition of children is limited to children who are the child of the head of household reported as living in the household in the Tanzanian 2012 census. I limit the analysis to include only households that have at least one child between age 7 to 17 years old who would be eligible for enrollment in school in order to prevent households without children or without eligible children from influencing the results[67]. This analysis is weighted using the household weights provided by the Tanzanian National Statistical Office to IPUMSI.

In both variations of the regression model, Table 3.10 shows that there is a decrease in the percent of children in the household enrolled in school if there is a child in the household with a disability. This main independent variable is modeled from the *binary* variable asking if a specific child has at least one sibling in the household with a disability. In Model 1, a household with a child with a disability has 7 percentage points lower school enrollment composition for all children, on average. In Model 2, there is no significant association between having a sibling with a disability in the household and the proportion of children without disabilities enrolled in schools. In fact, the direction is positive though very small and moderately significant ($\beta = 0.215$, $p=0.05$). The difference between Model 1 and Model 2 is the inclusion of children with disabilities in the numerator and denominator of the percent of children enrolled in school in Model 1 and only looking at children without disabilities in Model 2. As children with disabilities are less likely be enrolled in school, the results are intuitive. The inclusion of children with disabilities into the proportion of children enrolled in school in the household would reduce the overall proportion of children enrolled if the child with the disability is not enrolled.

As can be seen in Figure 3.6, the proportion of households in which all children are enrolled in school (proportion = 1, approximately 69 percent) is lower than in households in which at least one child has a disability (51 percent). However, when I exclude children with disabilities in the household, the proportion of children without disabilities in a household who are enrolled in school is 63 percent. This suggests that any difference between the proportion of

[67] Of households that meet these criteria, 2.75 percent (n=11,842; weighted n=113,428) have at least one child with a disability in the household.

eligible children enrolled in school in a household with and without children with disabilities is driven primarily by the lower enrollment of the children with disabilities themselves. The similarities between children without disabilities in households with no children with disabilities compared to their peers without disabilities but with a sibling with a disability shows similar results to the logistic regression analyses described above for the individual children.

Additional control variables included in the regression models are similar across the two specifications. Having a higher proportion of females in the household (related and unrelated) marginally increases the percent of children in the household enrolled in school. Each additional child of the head of household increases the proportion of children enrolled in school by 0.03 and 0.5 percentage points; this is likely due to the logic that the more children a family has, the higher the likelihood that at least one of them is enrolled in school at the time of the census. The total number of people in the household decreases the proportion of children enrolled in school. Finally, wealthier families and families in urban areas have, on average, a higher percent of children enrolled in school than poorer or rural families.

A final household-level analysis tests if having multiple children with disabilities in the household is associated with a decrease in the proportion of eligible children enrolled in school. Table 3.11 shows the results from a new independent variable that indicates if the household has one child with a disability or multiple children with disabilities as compared to households without any children with disabilities. This is based on the independent variable in the child level analysis that show the *categorical quantity* of children in the household and is different from the results in Table 3.10 (which includes the *binary* measure of if the household has at least one child with a disability). Model 1 shows the results for all school eligible children with and without disabilities and Model 2 shows the results for only children without disabilities. The result shown in Model 1 of Table 3.11 support the interpretation that if one child in the family has a disability, the percent of children enrolled in school decreased by 7.3 percentage points on average. For families with multiple children with disabilities in the family, the decrease in percent of children enrolled in school is 5.5 percentage points. Both of these results are significant. For children without disabilities (Model 2), having one sibling with a disability in the household increases the percent of children without disabilities enrolled in school by a significant, though small, 0.5 percentage points (p-value = 0.007) while having multiple siblings with disabilities decreases the percent of school enrolled children without disabilities by 2.2 percentage points (β =-2.28; p>0.001).

This analysis at the household level mirrors findings conducted with the child as the unit of analysis: they suggest there is not a significant effect on school enrollment for children without disabilities of having a sibling with a disability. The negative effect on the overall proportion of school enrollment among children 7 to 17 years old in the household with and without disabilities is largely because children without disabilities are less likely to be enrolled in school [as shown in the appendix and in other literature]. For children with multiple siblings with disabilities in the household, there is a slight negative effect on overall school enrollment for siblings without disabilities. Although when children were the main unit of analysis this result was not significant, the significant result in the household analysis can be partially attributed to the distribution of proportion of children without disabilities enrolled in school, as described above. A decrease of 2 percentage points in the proportion of school enrollment can be driven by households where one child not enrolled in school can influence the overall proportion by a large margin. For example, consider a household with four children all between ages 7 and 17 where two of the children have disabilities. The remaining two children without disabilities could have 100%, 50%, or 0% school enrollment in Model 2 of Table 3.11. These 50 percentage point changes between these three options can cumulate across all families in a way that influence the average statistics. Only 1.6 percent of households have a child with a disability (n=15,215 unweighted; n=146,493 weighted) while twice that many individual children have a sibling with a disability (n=39,506 unweighted; n= 378,878 weighted).

Sensitivity Checks on Robustness of Child-Level Logistic Regressions

In this next section, I return to the original logistic regression models presented in Table 3.8 and Table 3.9. The unit of analysis is individual children who do or do not have siblings with disabilities. First, I conduct analysis to further explore the wealth quintiles used in Table 3.9 as household wealth was a significant predictor of school enrollment. Next, I look at the results of school enrollment separately by gender. Finally, I construct two additional measures of school enrollment in primary and secondary school to explore the sensitivity of the main results on enrollment.

Wealth, school enrollment, and sibling status

The results of the main analyses show that wealth is a major factor in the likelihood that a child is enrolled in school. When household characteristics of urban status and wealth quintiles

are added to the model, sibling status is no longer a significant predictor of school enrollment compared to models without controls and with only demographic controls. In order to further explore the effect of wealth, I conduct sensitively tests to check the robustness of the created quintiles. First, I calculate deciles using the principal component analysis of household assets and housing characteristics. The results using deciles instead of quintiles produce identical coefficients across all operationalizations of sibling status.

When considering the family stress theory, the distribution of financial resources between children in a family where one or more of the children has a disability may differ according to wealth status. I tested the interaction between sibling status and wealth quintiles for all the quintiles. Compared to children in the same wealth quintile, there is no significant difference between children who have at least one sibling with a disability (binary operationalization) and children who do not. Additional tests using other operationalizations of sibling status find the same result. The coefficients on sibling status and quintiles (with the interactions) are identical to the model without the interaction. These results suggest that while wealthier children are more likely to be enrolled in school than children from less well-off families, this difference is not related to sibling status.

Gender of child and sibling status

The main results across all operationalizations of sibling status estimate that females have a slightly higher probability of being enrolled in school than males (OR 1.147), though the literature suggests a likelihood of female siblings being more likely to be caregivers and less likely to be enrolled in school (Evans 2010). Overall, I find that females are significantly more likely to be enrolled in school with and without household controls and across all ages, quintiles, family sizes, and urban status. These findings reflect the descriptive results in Table 3.6 that 80.75 percent of girls are enrolled in school compared to 78.16 percent of boys. There is no significance of the interaction between sibling status (across all operationalizations of sibling status) and gender of the siblings without a disability, as tested across all operationalizations of sibling status and the control variable for gender of reference child. The coefficients for the control variables are the same as in the main results.

I tested this further by dividing the analysis by sex of the child without disability. Table 3.12 shows the results by gender for three operationalizations of sibling status: *binary* indicator for having at least one sibling with a disability, *categorical composition* variable for having no, some, or all siblings with disabilities, and a *count* variable providing the number of siblings with disabilities a child has. The results suggest no significant difference in the odds of school

enrollment for female and male siblings without disabilities if they have a sibling with a disability compared to children of the same gender without a sibling with a disability.

The results presented in this chapter did not find a significant difference on school enrollment and sibling status by gender despite expected results related to gendered caregiving expectations. Some potential explanations for these results relate to recent policies by the Tanzanian government to increase education for female children (Wadsworth 2015). Additionally, other measures not adequately captured in census data, including children leaving school to work or earn money or children failing to pass the secondary school examination, are mitigating the effects of gender in this particular analysis. These hypotheses should be tested using alternative data sets that can identify sibling status and other factors that may contribute to school enrollment.

Primary School Students and Standard 7 Leavers

Older children are less likely to be enrolled in school across all models. As described in the paper, all Tanzanian children take an exam at the end of primary school, known as standard seven. Many students do not pass. For those who do enroll in secondary school, national curriculum switches from being taught in Swahili in primary school to English in secondary school. Additionally, there are far fewer secondary schools, and many require attendees to pay for boarding or private hostels near to the school. This is prohibitively expensive for many families. These challenges have created an entire cohort of young people known as "standard seven leavers"- those who stop attending school at the end of primary school. In 2016, for example, 70 percent of students passed the primary school exit examination, but only 33 percent of eligible children were enrolled in secondary school.

In order to identify the sensitively to my model towards these standard seven leavers, I ran several logistic regressions that identified if an individual was likely a standard seven leaver (they were not currently enrolled and their highest attained education level was standard seven or the end of primary). It may be hypothesized that siblings of persons with disabilities may be less likely to be enrolled in secondary school because of financial or familial responsibilities – as presented in the family stress theory—even if they passed the secondary school entrance exam. The results of this exam are not available in the census, though I use a variable that describes overall educational attainment to construct two model specifications. In Table 3.13, Model 1 includes only children whose maximum educational attainment is primary school. The sample includes children who are currently enrolled in primary school (and have not yet reached the

standard seven exam), children who are not enrolled in primary because they have dropped out of primary school, and those who never attended school. This allows me to look at the effect of sibling status on primary school enrollment only. Children enrolled in secondary school or who obtained at least some level of secondary school but are not currently enrolled are not included in this model. Model 2 creates a new independent variable that specifically identifies children likely to be standard seven leavers by creating an indicator for whether a child's highest level of education was standard seven, but they are no longer enrolled in school, compared to children who are currently enrolled at any level or who completed any level of secondary school.

In Table 3.13, the first column includes only individuals who are currently in primary school or who dropped out with less than a secondary education (Model 1). This is meant to capture children who are in enrolled who are in primary or are not enrolled, having dropped out after primary. The dependent variable in this analysis is current school enrollment. The findings of this analysis are almost identical to the main results.

The second column in Table 3.13 (Model 2) uses a new dependent variable that identifies the probability of being a standard seven leaver. The results in column 2 of Table 3.13 show that children are less likely to be standard seven leavers if they have at least one sibling with a disability (note the direction of the effect is reversed). This result is moderately significant (OR = 0.93, p = 0.1) and suggests that having a sibling with a disability may actually be slightly protecting. One possible explanation may be that the family may be able to invest more into the education of a child without disabilities if the sibling with disabilities is not able to attend school.

Tanzania 2002 Census

In order to test the robustness of both the classification of children with siblings with disabilities and school enrollment for children, I conduct parallel analysis using a sample of the 2002 Tanzanian census obtained from IPUMSI. The 2002 census differs from the 2012 census in the way questions about disabilities were asked. The 2002 census uses a more typical question format that asks, "Does this person have a disability?" and asks the respondent to indicate the disability instead of using the Washington Group disability questions. However, the data about disabilities in the 2002 and 2012 censuses are comparable in the IPUMSI census data because of the process used to create a binary variable comparable across time, as described in the data section above. The descriptive results showing the percent of children with disabilities is very similar in magnitude; children age 0-18 with a disability was 0.99 percent in 2002 compared to

1.07 percent in 2012. Among persons of all ages in 2002, 1.61 percent reported a disability compared to a similar 1.55 percent in 2012.

Despite the difference in question wording, the 2002 census yields very similar results to the 2012 census presented above, for both the child-level and household-level analysis. Using the Tanzania 2002 census data produces even less of an effect of sibling status on school enrollment than in the 2012 census. Across all operationalizations of sibling status, the odds ratios (without and with controls) for school enrollment for individual children are smaller and less significant than in 2012. Additionally, the analysis at the household shows that there is a slightly larger negative effect on having at least one sibling with a disability in the household (-11 percent compared to -7 percent in 2012), but this result remains insignificant. The similar results in 2002 and 2012 suggest that the methodology is robust to changes over time. However, the lack of improvement in the educational enrollment for siblings of children with disabilities across the decade shows the stagnation and insufficient efforts to support children with disabilities and their siblings.

3.6 Discussion

The two primary aims of this chapter were to establish of population estimates of children living with siblings with disabilities and measure the impact on sibling status on school enrollment for children with siblings with disabilities compared to children without siblings with disabilities. Based on the analysis of the data from the 2012 Tanzanian census, approximately 510,000 children (3.5 percent of children age 0-17 years-old) in Tanzania either have a disability or have a sibling with a disability. For every child with a disability (an estimated 170,000 children), there are approximately 2.5 additional brothers or sisters under the age of 18 who live with them and who do not have disabilities (approximately 340,000 siblings). The quantification of this population is important to convince policy makers that the scope and impact of childhood disability is larger than simply the number of young people with disabilities[68].

[68] The limitations of using census data suggest that this estimate is an under-estimate of the true value. The analysis sample includes only children under 18 who are living at home with a parent who is the head of household — a sub-set of children who might be experiencing a sibling or close relative with a disability. It also potentially excludes some of the most vulnerable children such as orphans, children not living with parents, or children missed in the census. Therefore, it is likely that the estimates are higher than those presented in this chapter.

The analysis of school enrollment presented in this chapter is situated within a theory of family systems that has historically focused on negative outcomes for individuals with a sibling with a disability. My hypothesis that children who have siblings with disabilities would be less likely to be enrolled in school falls within the influence of this older tradition and it adapts concepts of disability theory developed in Western contexts that may not be appropriate for Tanzania or other Sub-Saharan African countries. The results of the analysis suggest that there is no relationship between sibling status and school enrollment for children who have siblings with disabilities. This finding was robust regardless of the operationalization of sibling status. The null findings are important for the field of sibling research because they suggest that Western theories of sibling relationships and family structures may not be applicable to children in non-Western countries.

In conclusion, I connect the results of the analysis to three main points presented in the introduction. First, I argue that in order to explore various siblings relationships *within* a culture — such as the differences between siblings where one or more have a disability and siblings without disabilities within Tanzania — there needs to be a greater emphasis on large-scale, quantitative and demographic data. Second, the established literature on sibling relationships rarely accounts for cross-cultural differences (i.e. differences *between* cultures), instead retaining Western-centric ideals of childhood and interpretations of disability. Third, sibling relationships must be considered within the cultural context shaped by disability, education, and family policies. I will describe each of these in turn.

The lack of nationally-representative estimates of the number of children who have at least one sibling with a disability is a long-identified methodological challenge in disability research (Stoneman 2005). My analysis can be a starting point for other large-scale quantitative research on siblings in Tanzania and Sub-Saharan Africa. The methodology for identifying potential siblings presented in this chapter can be expanded to other countries and other large, nationally-representative datasets that allow for more dependent variables to be explored and to test the robustness of the proposed methodological operationalization of sibling status.

Knowing more about families where one or multiple of children have a disability can help identify specific needs that families, siblings, and persons with disabilities may have. Comparative studies of sibling relationships can utilize a variety of theoretical frameworks, such as the family structure, family stress, or life course theories in ways that holistically represent the interrelationships and long-term experiences of sibling of persons with disabilities. The results presented in this chapter are the first step in knowing more about the nearly half million children with disabilities or siblings with disabilities in Tanzania. Census data can only provide high-level

interpretations of results on just one relationship: sibling status and current school enrollment. Using school enrollment, as opposed to literacy, school attendance, or educational attainment, reflects the experiences of the siblings at a moment in time and is a crude measure of education[69]. This measure can only be understood to represent the cross-sectional data collected at the moment of the census.

In contrast, longitudinal data would allow additional understanding of the patterns of schooling for siblings of persons with disabilities. This type of research would incorporate the family stress theory of different life events such as having additional children or the later change in disability status due to an illness or accident. Longitudinal data would also allow research on sibling relationships to be situated in life course theory, as it is well established that sibling relationships change across the life course. In these situations, using measures of educational grade completion or timing of school enrollment across different ages or genders would allow for a larger picture of the relationship between having a sibling with a disability and overall educational attainment. Census data can provide only a snapshot of a moment; measuring current school enrollment is only the first step in understanding the situation of siblings in Tanzania in 2012. Beyond the results concerning school enrollment presented in this chapter, future research using other sources of nationally-representative data[70] could incorporate additional variables on health, service utilization, family status and composition, geography, and more detailed education information such as examples just described. Using large-scale demographic data continues to push the field of sibling research beyond studies that commonly use small-scale or convenience sampling.

While large-scale data can continue to push the field to look at different sibling relationships *within* cultures, there is still major work to be done on considerations of cross-cultural theoretical interpretations of disability and sibling relationships. The findings in this chapter highlight a need for continued improvement of statistics on disability status. The use of the Washington Group variables in the Tanzanian census is one way that global scholars in the field of disability are continuing to expand and integrate culturally diverse interpretations and definitions of disability beyond the medical model developed in a Western context. Using the Washington Group of categorizations of "a lot of difficulty" or are "completely unable" to create a binary indicator of "disabled" only captures a portion of children who are experiencing

[69] There are critiques of using measures of both educational attainment and literacy (Smith-Greenaway 2015)

[70] Example of such nationally-representative surveys include the Demographic and Health Surveys (DHS) or the Multiple Indicator Cluster Surveys (MICS) commonly and frequently collected in many developing countries.

difficulties or children who have a sibling who might need additional support to perform activities. As Table 3.4 shows, there are more children who were reported as experiencing "some difficulty" in one or more of the activities compared to children reported as having "a lot of difficulty" or "unable to do X". Therefore, the findings in the analysis may underestimate the effects of having a sibling with a disability compared to if children with "some" difficulty were additionally included in the analysis. More research needs to be done about how the Washington Group questions are understood and interpreted in multiple African and Tanzanian cultures.

Additionally, the results about school enrollment that I present do not and cannot take into account important aspects of Tanzanian culture. The research presented here on Tanzanian educational outcomes is situated within the family systems theories. However, the findings presented can begin to situate sibling relationships in Tanzania into a bioecological framework or sibling embedded systems framework developed specifically for Tanzania. Findings about educational enrollment for siblings of children with disabilities (a microsystem finding) can be further explored to understand how sibling relationships change educational outcomes across the life course (chronosystem). The interactions between families and the school systems, and other support systems can identify areas where the child with disabilities is being assisted appropriately and thus the sibling relationship may be influenced at the exosystem. Situating Tanzanian families within the broader educational policies like the National Strategy for Inclusive Education and the disability policies such as the National Policy on Disability (2004) and the Persons with Disabilities Act (2010) identifies how the macrosystems impact individual families. Policies aimed at additional supports for families of children with disabilities may provide children with fewer caregiving responsibilities or more financial support to pay for educational expenses (family stress theory).

Finally, it is essential for policy makers to recognize how complexities of different types of sibling relationships within a specific culture and differences between cultures shape the policies that support persons with disabilities and their families. Policies that support persons with disabilities must recognize the relationship that persons with disabilities have with their families and siblings in order to be successful. Siblings are major players in the child's life and the siblings with disabilities greatly influence non-disabled siblings' lives. These interactions and experiences are both negative and positive. Siblings are on the front lines of caregiving. I argue that paying attention to siblings does not take attention away from persons with disabilities and their needs. In fact, supporting persons with disabilities takes the entire family and entire community. Drawing attention to siblings of persons with disabilities on a national scale can bring attention and focus to an otherwise hidden population. Therefore, supporting siblings is an

additional way to continue to move forward on inclusive and progressive policies aimed to improve the lives of children and adults with disabilities.

3.7 Figures

Figure 3.1: Population Pyramid of Age (in years) of Children of the Head of Household [Weighted]

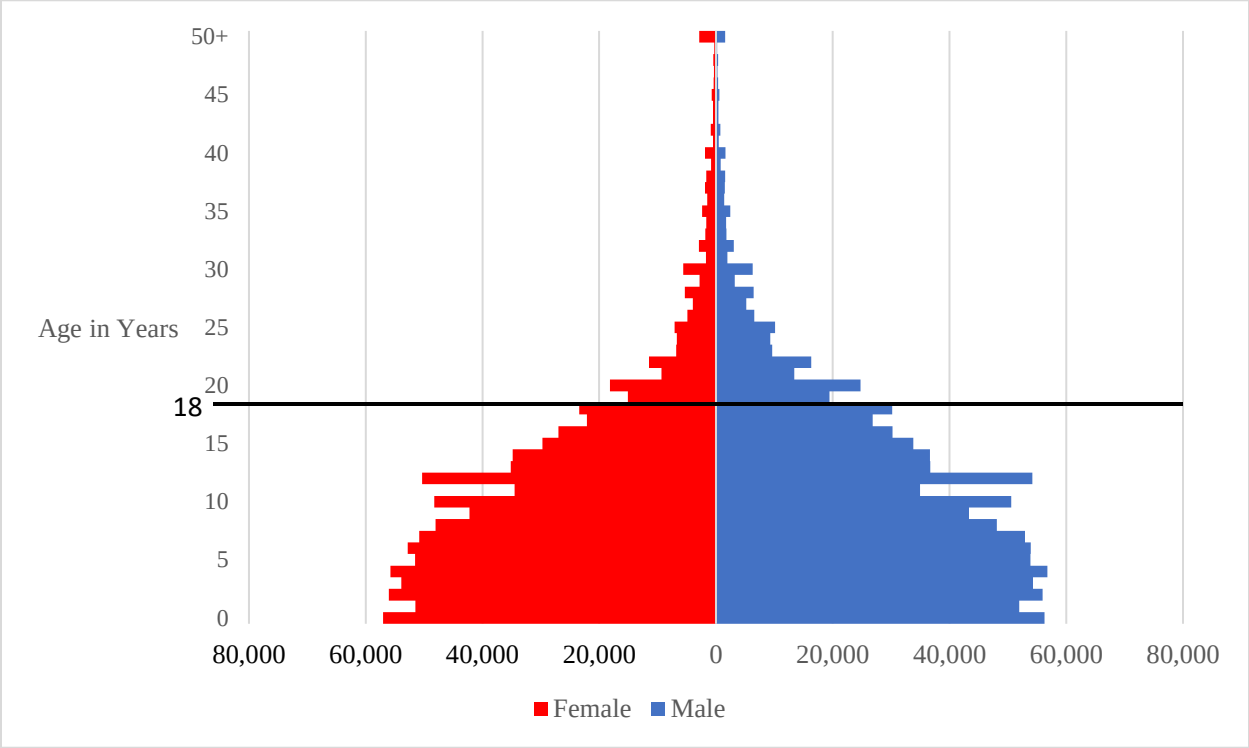


Figure 3.2: Percent of children (0-17) with 1 or 2+ Siblings with a disability (by disability status) [Weighted]

Figure 3.3: Distribution of the Categorical Composition of sibling for each child (0-17) by disability status [Weighted]

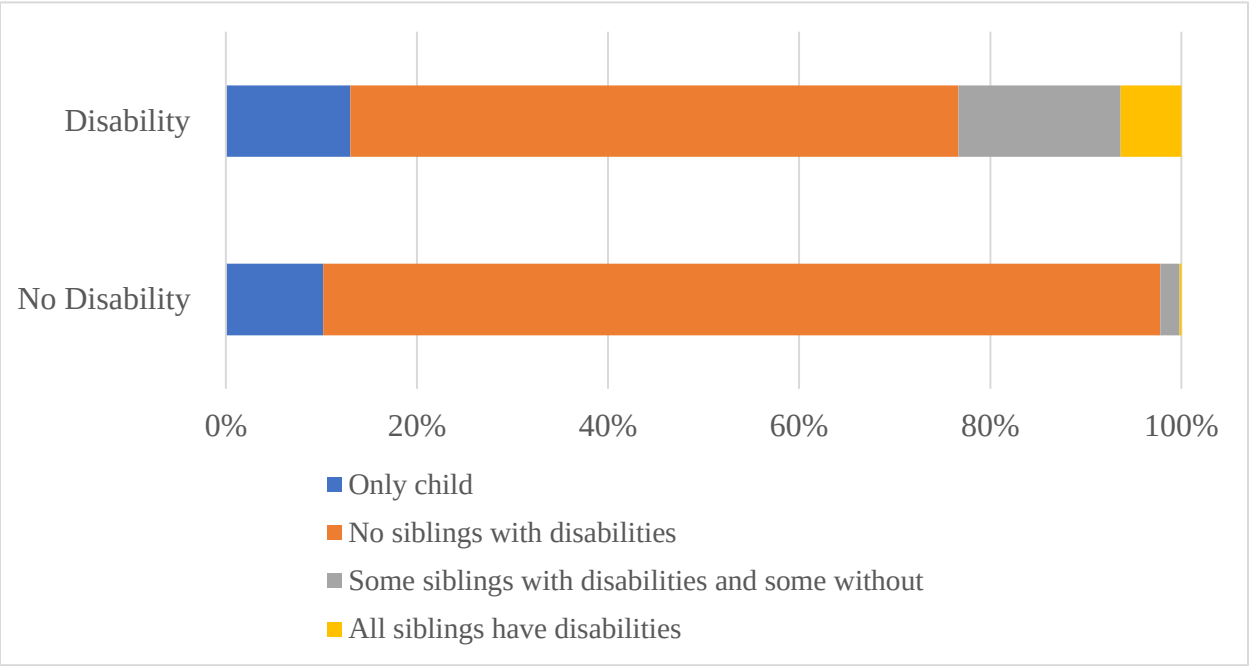
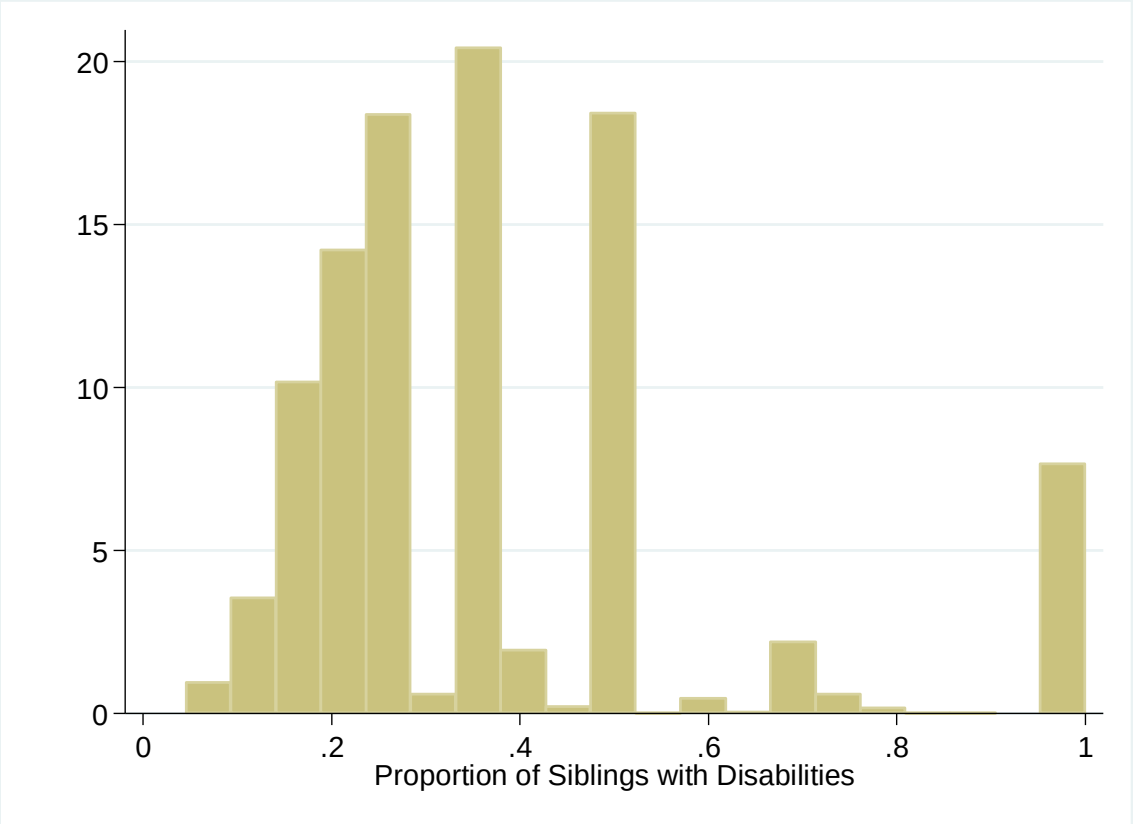


Figure 3.4: Proportion of siblings with disability that each child has for children with at least one sibling with a disability.



Tallest bars at .25, .33. and .5.

Figure 3.5: School enrollment for children 7-17 years old by disability status and sibling composition

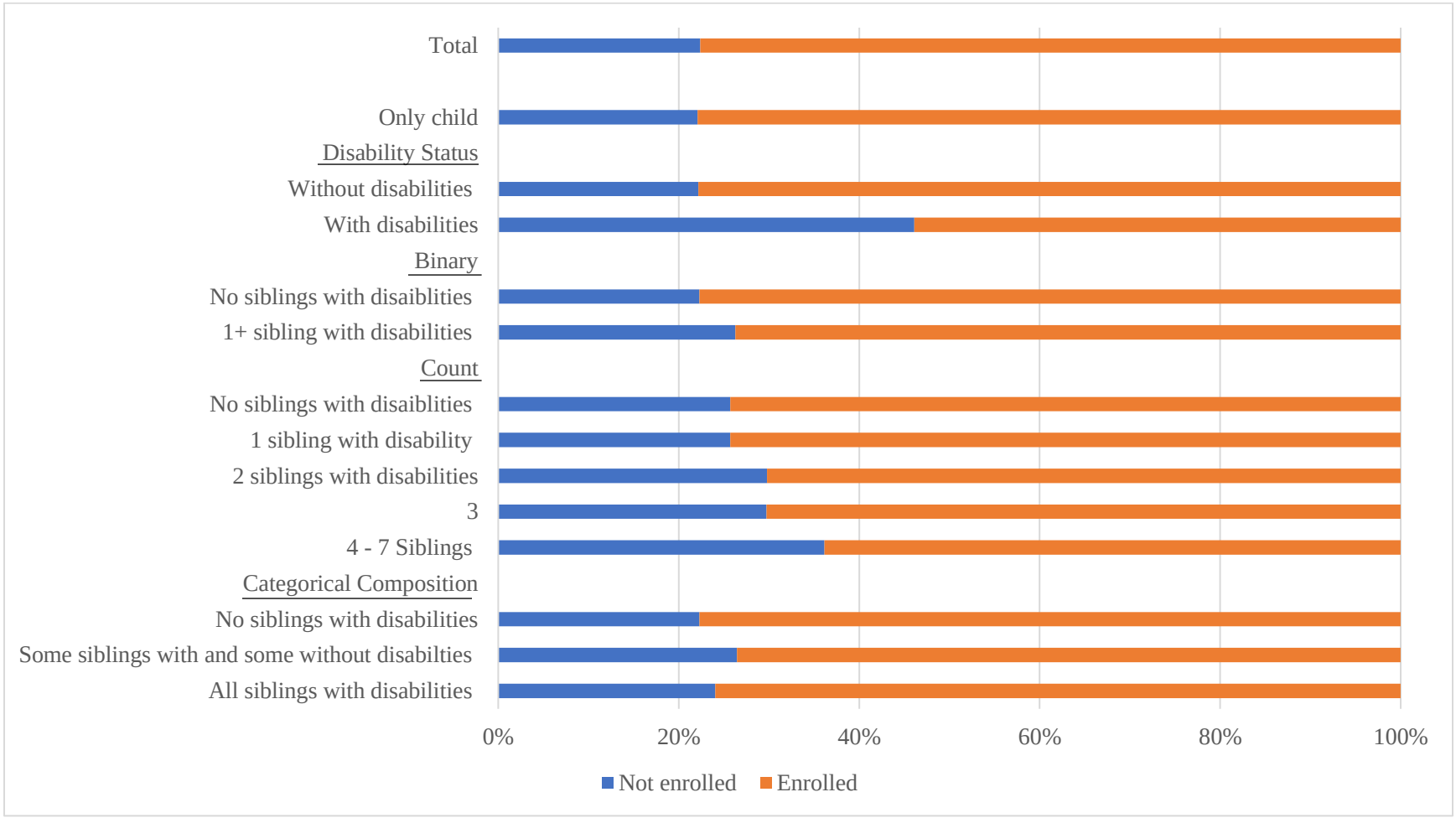
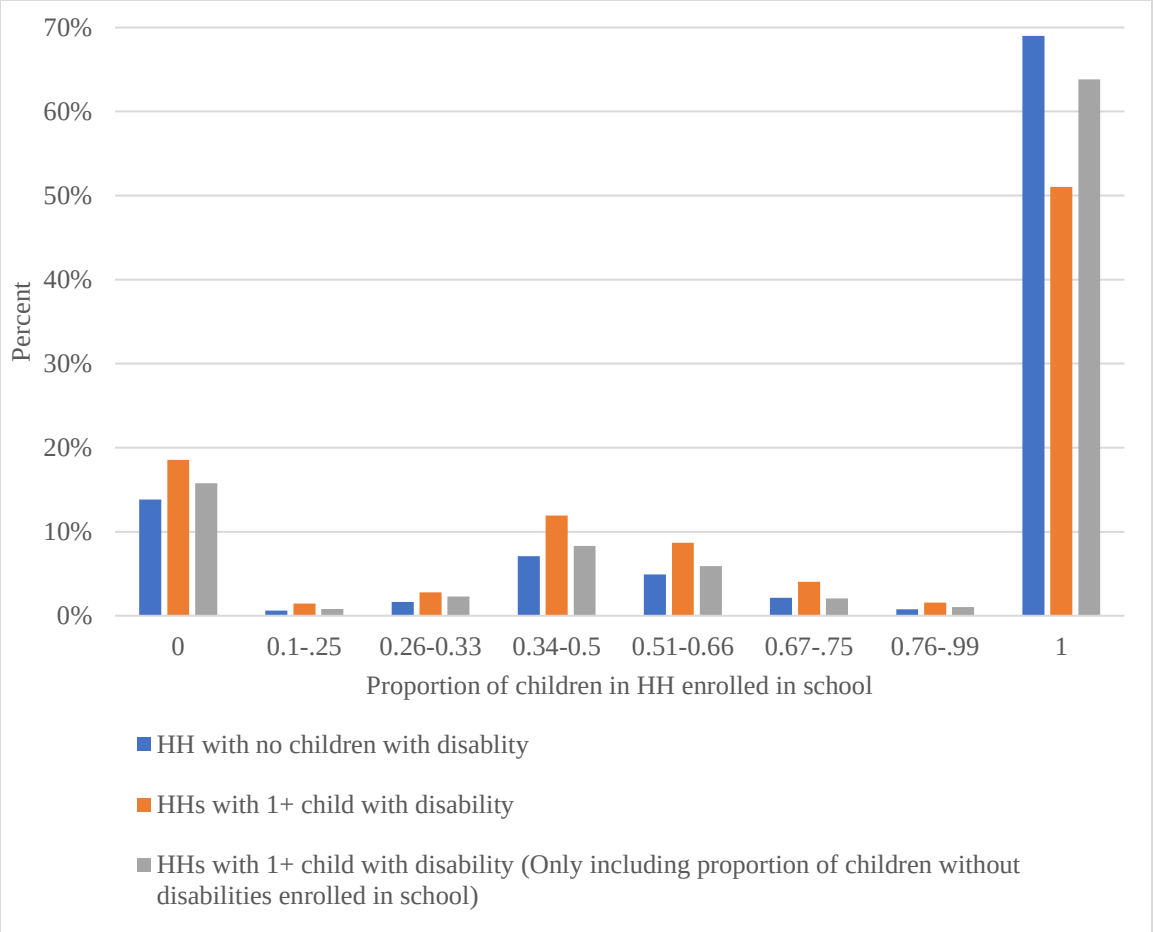


Figure 3.6: Histograms of the proportion of children enrolled in school for (1) households with no children with disabilities, (2) children with and without disabilities in households with at least one child with a disability, and (3) children without disabilities in households with at least one child with a disability



3.8 Tables

Table 3.2: Operationalization of Sibling Status

Operationalized Variable	Definition	Variable Type	Expression	Shorthand in analysis tables
Binary	Child has at least one sibling with a disability	Categorical	1= Child has 1+ sibling with disability 0= otherwise	0/1+
QUANTITY				
Count	Number of siblings with disabilities	Continuous numeric	Data includes children with between 0 -7 siblings with disabilities	*Count* (if treated as dummy indicators for each number) *Cont.* (if treated a continuous numeric)
Categorical quantity	No siblings, one sibling, or multiple siblings (2+) with disabilities	Categorical	0 = No siblings with disabilities 1 = 1 sibling with disabilities 2 = 2+ siblings with disabilities	0/1/2+ 0/1/2/3+
COMPOSITION				
Categorical composition	None, some, or all siblings with disabilities	Categorical	0 = No siblings with disabilities 1 = Some siblings have disabilities, and some do not (regardless of number of total siblings) 3 = All siblings have disabilities (regardless of number of total siblings)	*Comp.*
Proportion	Proportion of the number of siblings with disabilities out of the total siblings	Proportion	0 (no siblings with disabilities) to 1 (all siblings with disabilities)	*Prop*

Table 3.3: Example households showing categorization of disability and sibling status across different operationalizations

		Binary		**Quantity**		**Composition**	
	Disability Status	*Binary Category* (Table 3.1)	*Binary*	*Count*	*Categorical Quantity*	*Categorical Composition*	*Proportion*
Household 1							
Child 1	No	A	No	0	0	None	0
Child 2	No	A	No	0	0	None	0
Household 2							
Child 1	No	B	Yes	1	1	Some	1/2
Child 2	No	B	Yes	1	1	Some	1/2
Child 3	Yes	C	No	0	0	None	0
Household 3							
Child 1	No	B	Yes	2	2	Some	2/3
Child 2	No	B	Yes	2	2	Some	2/3
Child 3	Yes	D	Yes	1	1	Some	1/3
Child 4	Yes	D	Yes	1	1	Some	1/3
Household 4							
Child 1	No	B	Yes	1	1	All	1
Child 2	Yes	C	No	0	0	None	0

Table 3.4: Disability status by Washington Group Characteristic by type of disability and severity for children of the head of household, age 0-17

	Seeing	Hearing	Walking	Remembering	Self-care
No difficulty	99.51%	99.40%	99.57%	99.51%	90.96%
Some difficulty	0.40%	0.45%	0.24%	0.28%	0.29%
A lot of difficulty	0.07%	0.10%	0.08%	0.09%	0.10%
Unable	0.02%	0.06%	0.11%	0.12%	0.47%
Not applicable					8.19%

Note: The first four columns in table 3.4 are comparable to one another [Weighted; n= 1,630,583]. The final column labeled "self-care" has a different universe of children reporting difficulty with self-care. The census enumeration instructions state that children ages 0-3 be coded as "Not applicable", however some persons reporting to be between 0 and 3 years old as indicate a difficulty level with self-care. The sample size in this final column is 1,500,608.

Table 3.5: Number of 0-17 year-old children of heads of household living at home by disability status, for four operationalizations of sibling status

		No Disability			Disability			Total		
		Pop. Est.	Pop (%)	Sample	Pop. Est.	Pop (%)	Sample	Pop. Est.	Pop (%)	Sample
	Total	15,669,641	100%	1,613,096	170,078	100%	17,488	15,839,718	100%	1,630,584
Reference children with …										
Only Children	No siblings	1,702,636	11%	164,157	23,109	14%	2,279	1,725,745	11%	166,436
	With siblings	13,967,005	89%	1,613,096	146,969	86%	15,209	14,113,973	89%	1,464,148
Binary	No sib. with dis.	13,626,974	87%	1,413,516	108,122	64%	11,126	13,735,096	87%	1,424,642
	1+ sib. with dis.	340,031	2%	35,423	38,847	23%	4,083	378,878	2%	39,506
Count	No sib. with dis.	13,626,974	87%	1,413,516	108,122	64%	11,126	13,735,096	87%	1,424,642
	1 sibling with dis.	306,251	2%	31,855	28,050	16%	2,930	334,301	2%	34,785
	2	27,580	0.18%	2,911	7,519	4%	792	35,099	0.20%	3,703
	3	4,995	0.03%	528	1,894	1.10%	212	6,889	0.04%	740
	4	677	0.00%	74	885	0.50%	100	1,562	0.01%	174
	5	404	0.00%	42	418	0.20%	42	822	0.01%	84
	6	112	0.00%	12	82	0.05%	7	194	0.00%	19
	7	12	0.00%	1	0	0.00%	0	12	0.00%	1
Cat. Quantity	No sib. with dis.	13,626,974	87%	1,413,516	108,122	64%	11,126	13,735,096	87%	1,424,642
	1	306,251	2%	31,855	131,231	77%	13,405	437,482	3%	45,260
	2+	33,780	0.20%	3,568	38,847	23%	4,083	72,626	0.50%	7,651
Cat. Comp.	No sib. with dis.	13,626,974	87%	1,413,516	108,122	64%	11,126	13,735,096	87%	1,424,642
	Mixed	310,109	2%	32,400	28,345	17%	2,972	338,454	2%	35,372
	All sibs. have dis.	29,922	0.20%	3,023	10,502	6%	1,111	40,424	0.30%	4,134

Table 3.6: Table of demographic and household control variables by school enrollment and disability status.

	Total			**Without disabilities**			**With disabilities**		
	Not enrolled	Currently enrolled	Total	Not enrolled	Currently enrolled	Total	Not enrolled	Currently enrolled	Total
Age (mean)	11.67	11.30	11.38	11.66	11.30	11.38	11.98	11.43	11.68
	(3.590)	(2.862)	(3.044)	(3.595)	(2.862)	(3.044)	(3.264)	(2.849)	(3.059)
Prop. female	0.455	0.495	0.486	0.456	0.495	0.487	0.439	0.460	0.451
	(0.498)	(0.500)	(0.500)	(0.498)	(0.500)	(0.500)	(0.496)	(0.498)	(0.498)
# children in HH	4.318	3.822	3.933	4.326	3.823	3.935	3.823	3.761	3.789
	(2.277)	(1.889)	(1.993)	(2.279)	(1.889)	(1.993)	(2.056)	(1.870)	(1.958)
Urban	0.110	0.319	0.272	0.109	0.319	0.273	0.181	0.289	0.239
	(0.313)	(0.466)	(0.445)	(0.312)	(0.466)	(0.445)	(0.385)	(0.453)	(0.427)
Wealth Quintiles									
1 (Lowest)	0.377	0.167	0.214	0.378	0.167	0.214	0.332	0.203	0.263
	(0.485)	(0.373)	(0.410)	(0.485)	(0.373)	(0.410)	(0.471)	(0.403)	(0.440)
2	0.277	0.189	0.209	0.278	0.189	0.209	0.237	0.202	0.218
	(0.448)	(0.391)	(0.406)	(0.448)	(0.391)	(0.406)	(0.425)	(0.401)	(0.413)
3	0.183	0.211	0.205	0.183	0.211	0.205	0.190	0.215	0.203
	(0.387)	(0.408)	(0.404)	(0.387)	(0.408)	(0.404)	(0.392)	(0.411)	(0.403)
4	0.126	0.232	0.208	0.125	0.232	0.208	0.165	0.211	0.190
	(0.332)	(0.422)	(0.406)	(0.331)	(0.422)	(0.406)	(0.371)	(0.408)	(0.392)
5 (Highest)	0.0370	0.200	0.164	0.0363	0.201	0.164	0.0759	0.169	0.126
	(0.189)	(0.400)	(0.370)	(0.187)	(0.400)	(0.370)	(0.265)	(0.375)	(0.332)
N	197,511	684,788	882299	194,086	680,783	874869	3,425	4,005	7430

For children (of the head of household) age 7-17 (both with and without disabilities) [Weighted]
Note: Never enrolled includes children enrolled in the past and children never enrolled. % or mean (SD in parenthesis)

Table 3.7: Odds of school enrollment for children with at least one sibling with a disability (*binary* variable)

	(I)	(II)	(III)
	Basic	Demographic	All controls
Sibling(s) with disability (ref. no siblings with disabilities)	0.81***	0.93***	1.01
	(0.02)	(0.02)	(0.02)
Age		0.97***	0.95***
		(0.00)	(0.00)
Female		1.16***	1.15***
		(0.01)	(0.01)
N children in HH <18		0.86***	0.91***
		(0.00)	(0.00)
Wealth Quintile (Ref. lowest)			
2			1.55***
			(0.02)
3			2.59***
			(0.03)
4			3.79***
			(0.04)
5 (highest)			8.83***
			(0.18)
Urban			1.85***
			(0.02)
N	803745	803745	803745

Notes: Clustered robust standard errors at the household in parenthesis. *p<0.10 **p< 0.05 *** p<0.010

Table 3.8: Odds of being enrolled in school for seven different of operationalization without control variables (ref. group is children who have no siblings with disabilities)

	I	II	III	IV	V	VI	VII
	0/1+	*0/1/2+*	*0/1/2/3+*	*Contin*	*Count*	*Comp.*	*Prop*
1+ Sib. w/ dis.	0.81***						
	(0.02)						
1 Sib. w/ dis.		0.84***					
		(0.02)					
2+ Sib. w/ dis.		0.65***					
		(0.04)					
1 Sib. w/ dis.			0.84***				
			(0.02)				
2 Sib. w/ dis.			0.69***				
			(0.05)				
3+ Sib. w/ dis.			0.537***				
			(0.08)				
# of Sib. w/ dis.				0.83***			
				(0.01)			
# Sib. w/ dis.: 1					0.84***		
					(0.02)		
2					0.69***		
					(0.05)		
3					0.59**		
					(0.09)		
4					0.44**		
					(0.14)		
5					0.16**		
					(0.14)		
6+					1.12		
					(0.78)		
Some sib. w/ dis.						0.80***	
						(0.02)	
All sib. w/ dis.						0.97	
						(0.06)	
Prop. of sib. in HH w/ dis.							0.62***
							(0.04)
N	803745	803745	803745	803745	803745	803745	803745

Notes: Clustered robust standard errors at the household in parenthesis. *p<0.10 **p< 0.05 *** p<0.010

Table 3.9: Odds Ratios for enrollment in school for different operationalizations of sibling (with controls)

	I *0/1+*	II *0/1/2+*	III *0/1/2/3+*	IV *Contin*	V *Count*	VI *Comp.*	VII *Prop*
1+ Sibling with disability	1.01 (0.02)						
1 Sibling with disability		1.02 (0.02)					
2+ Siblings with disabilities		0.94 (0.06)					
1 Sibling with disability			1.02 (0.02)				
2 Siblings with disabilities			0.95 (0.07)				
3+ Siblings with disabilities			0.69 (0.25)				
# of Sibling with disability				1.00 (0.02)			
# Siblings with disabilities: 1					1.02 (0.02)		
2					0.95 (0.07)		
3					1.00 (0.16)		
4					0.70 (0.29)		
5					0.45 (0.35)		
6+					2.28 (2.07)		
Some sibs. w/ dis.						1.03 (0.02)	
All siblings with disabilities						0.75*** (0.05)	
Prop. of sibs. in HH w/ dis.							0.93 (0.05)
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	803745	803745	803745	803745	803745	803745	803745

Notes: Demographic controls: Sex, age, number of children in household. Household controls: Wealth quintile, urban/rural status. Clustered robust standard errors at the household in parenthesis. *p<0.10 **p< 0.05 *** p<0.010

Table 3.10: Regression of percent of children enrolled in school in a household for households with children eligible to be enrolled in school (having a child between 7-17) using a binary indicator for having at least one child in the household with a disability.

	(1) Percent of household's children (7-17) currently enrolled in school	(2) Percent of household's children (7-17) without *disabilities* currently enrolled in school
Child with dis. in household	-7.060***	0.215**
	(0.10)	(0.11)
% females in household	0.038***	0.038***
	(0.00)	(0.00)
N own children in household	0.452***	0.374***
	(0.01)	(0.01)
N persons in household	-0.922***	-0.979***
	(0.01)	(0.01)
Quintiles (Lowest = 1)		
2	10.230***	10.228***
	(0.05)	(0.05)
3	19.195***	19.249***
	(0.05)	(0.05)
4	24.568***	24.585***
	(0.05)	(0.05)
5 (Highest)	29.762***	29.745***
	(0.06)	(0.06)
Urban = 1	6.566***	6.562***
	(0.05)	(0.05)
Constant	62.564***	62.788***
	(0.07)	(0.07)
N	4,136,106	4,119,486
R^2	0.124	0.123

Note: Clustered robust standard errors at the household in parenthesis. *p<0.10 **p< 0.05 *** p<0.010

Table 3.11: Regression of percent of children enrolled in school in a household for households with children eligible to be enrolled in school (having a child between 7-17) using a categorical quantity variable for having 1 or multiple children in the household with disabilities.

	(1) Percent of household's children (7-17) currently enrolled in school	(2) Percent of household's children (7-17) without *disabilities* currently enrolled in school
1 Child with disability in HH	-7.259***	0.536***
	(0.11)	(0.12)
Multiple (2+) children with disabilities in HH	-5.545***	-2.282***
	(0.29)	(0.32)
% females in household	0.038***	0.038***
	(0.00)	(0.00)
N own children in household	0.451***	0.376***
	(0.01)	(0.01)
N persons in household	-0.980***	-0.979***
	(0.01)	(0.01)
Quintiles (Lowest = 1)		
2	10.229***	10.228***
	(0.05)	(0.05)
3	19.196***	19.248***
	(0.05)	(0.05)
4	24.570***	24.584***
	(0.05)	(0.05)
5 (Highest)	29.762***	29.745***
	(0.06)	(0.06)
Urban = 1	6.566***	6.563***
	(0.05)	(0.05)
Constant	62.569***	62.780***
	(0.07)	(0.07)
N	4,136,106	4,119,486
R^2	0.124	0.123

Note: *p<0.10 **p< 0.05 *** p<0.010

Table 3.12: Odds Ratios for school enrollment by gender and three operationalizations of sibling status

	(1) Binary girls	(2) Binary boys	(3) Cat girls	(4) Cat boys	(5) Count girls	(6) Count boys
1+ Sibling with disability	1.01	1.01				
	(0.03)	(0.03)				
Some Sibling with disability			1.04	1.03		
			(0.03)	(0.03)		
All Sibling with disability			0.71***	0.79**		
			(0.07)	(0.08)		
# of Sibling with disability					1.00	1.00
					(0.02)	(0.02)
Age	0.94***	0.96***	0.94***	0.96***	0.94***	0.96***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Quintile (ref. lowest)						
2	1.56***	1.55***	1.56***	1.55***	1.56***	1.55***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
3	2.72***	2.49***	2.72***	2.49***	2.72***	2.49***
	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
4	3.97***	3.65***	3.97***	3.65***	3.97***	3.65***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
5 (Highest)	8.27***	9.46***	8.27***	9.46***	8.27***	9.46***
	(0.22)	(0.25)	(0.22)	(0.25)	(0.22)	(0.25)
Urban = 1	1.77***	1.91***	1.77***	1.91***	1.77***	1.91***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Total sib. in HH	0.90***	0.91***	0.90***	0.91***	0.90***	0.91***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	391966	411779	391966	411779	391966	411779

Note: *p<0.10 **p< 0.05 *** p<0.010

Table 3.13: Odds Ratios of Logistic Regression on Current Primary School Enrollment and Standard Seven Leavers and Sibling Status

	(1) Dep. Variable: Current Primary School Enrollment	(2) Dep. Variable: Std7 Leaving
1+ Sibling with disability	1.02	0.93*
	(0.02)	(0.04)
Age	0.92***	2.32***
	(0.00)	(0.01)
Sex (ref. Female)	1.13***	1.15***
	(0.01)	(0.02)
Quantiles		
(Ref. = Lowest)		
2	1.54***	0.87***
	(0.01)	(0.02)
3	2.51***	0.66***
	(0.03)	(0.01)
4	3.59***	0.51***
	(0.04)	(0.01)
5 (Highest)	7.83***	0.20***
	(0.16)	(0.01)
Urban status	1.79***	0.59***
	(0.02)	(0.01)
Total siblings in HH <18	0.91***	1.03***
	(0.00)	(0.00)
N	745,914	662,398

Note: *p<0.10 **p< 0.05 *** p<0.010

Chapter 4:

The Impact of the 2014-2015 West African Ebola Outbreak on Fertility in Sierra Leone

Abstract:

This study examines the social impacts of changes to fertility patterns in Sierra Leone before and after the 2014 West African Ebola outbreak. Short-term fertility patterns may have been altered due to behavior modifications encouraged by public health officials during this time of shock. During the outbreak and the state of emergency, people were encouraged to limit social interactions and physical contact with others to avoid Ebola transmission through contact with bodily fluids. Such social changes at the societal level may alter both aggregate and individual fertility decisions. Using data from a nationally representative sample of women in Sierra Leone, I estimate the effect of the Ebola outbreak as a reduction to fertility levels at the district level using difference-in-difference models. I also hypothesize individual changes to the timing of births using event history analysis models. I find that fertility at the district level decreased during the outbreak by 0.4 births per district per month, and the length of the birth interval for individual women increased by 16%. This suggests that fertility decisions vary in the short term and the social impact of Ebola has been underemphasized.

4.1 Introduction

The worst recorded outbreak of Ebola Virus Disease (EVD or Ebola) spread through West Africa in late 2014 and through 2016. The three main countries affected by the outbreak — Guinea, Liberia, and Sierra Leone — attributed approximately 29,000 cases and 11,000 deaths to Ebola (World Health Organization (WHO) 2016). Sierra Leone reported the highest number of cases — 14,124 cases and 3,956 reported deaths across the country. The scope and scale of the 2014 outbreak led to medical and epidemiological advances for those infected and survivors. The long-term social consequences of reduced health care utilization and reproductive health services during and after the outbreak will have lasting demographic effects on population dynamics due

to excess mortality and morbidity, reduced life expectancy, and changes to fertility patterns and behaviors (Lévy et al. 2018; Rohwerder 2014; Elston et al. 2017)[71]. This study uses data from Sierra Leone to examine the impacts of Ebola to population fertility patterns including total numbers of children and timing of births before, during, and after the 2014 Ebola epidemic via potential behavioral and biological mechanisms.

The theoretical and empirical impacts of the West African Ebola outbreak in Sierra Leone on fertility are both biological and social. The biological effects of Ebola on fertility are potentially serious. Medical and epidemiological research suggests that contracting Ebola while pregnant results in higher miscarriage rates along with other negative outcomes for in-utero exposures and fetal development; there is a 100% recorded mortality rate for neonates (Fallah et al. 2016; Jamieson et al. 2014; Iliyasu, Dattijo, and Habib 2017; Godwin et al. 2019). There is no evidence that pregnant women have a higher risk of contracting Ebola or higher mortality if they contract Ebola. Another biological complication to reproduction is the potential for Ebola to be sexually transmitted in the sperm of male survivors (Mate et al. 2015; Deen et al. 2015). These biological findings, while somewhat limited in evidence, suggest possible long-term impacts on fertility based on the fecundity of survivors and women who did not contract the virus.

Clinical publications about biological risks to fertility were not reported or widely publicized until after the start of the public health crisis in West Africa. But enough was known about the contagiousness and deadliness of Ebola to promote swift public health campaigns to spread messages of Ebola prevention and risk reduction measures over television, radio, and billboards (Gidigo et al. 2015). These messages reached people throughout Sierra Leone, even in areas of low exposure and lower risk of contracting Ebola. People were encouraged to avoid physical contact with others as Ebola is transmitted most easily through contact with bodily fluids. These social changes could have affected the timing of fertility by reducing sexual activity and by limiting opportunities for socialization when potentially-infected people were quarantined (Fairhead 2014). An anthropological study published by Lipton (2019) describes a pregnant couple in Freetown during the Ebola epidemic and their challenges accessing pre and post-natal care, restrictions of movement that impact employment, cancellation of traditional celebrations, and limitations of familial support. Hearing about similar experiences of couples across the

[71] The terms outbreak and epidemic can be used interchangeably when talking about the situation in 2014-2016 West Africa (Oran 2018).

country may have led to changes in large scale fertility patterns as couples may have explicitly decided or unintentionally been compelled to delay or forgo childbearing.

Building upon these qualitative examples, I use data from the Sierra Leone 2016 Demographic and Health Survey to look at nationally representative patterns of fertility change in Sierra Leone. I employ a study design that compares the effect of the Ebola outbreak on districts with Ebola to districts not exposed to Ebola through May 2014 through April 2015, the main peak of the outbreak. Results suggest that the Ebola outbreak contributed to a short-term reduction in fertility, though due to data limitations I cannot directly test specific biological or behavioral mechanisms, nor can I identify survivors of Ebola directly. This analysis explores both changes in fertility across the population and changes in individual women's fertility.

High mortality and thus relatively few survivors resulted in little being known about the long-term health impacts among survivors of earlier and smaller-scale outbreaks of Ebola. In the 2014 West African outbreak, there were many more survivors across a greater geographic area. Despite the global coverage of the virus spread, Ebola only accounted for 14 percent of deaths in Sierra Leone in 2015, behind malaria at 27 percent (Seisay and Kamara 2017). The vast majority of the population was not infected. Based on population figures and reported Ebola cases and death, the infection rate was 1.2 to 1.9 infected persons per thousand persons across the general population and the death toll of 3,956 implies 5.5 deaths per 10,000 persons(World Health Organization 2016). Any large-scale changes to fertility patterns are more likely to be due to people reacting to the epidemic behaviorally than directly due to biological effects of the virus itself. Population-level changes in fertility during and after the 2014-2016 Ebola outbreak can suggest how the presence of Ebola in a district, and the assumed public health awareness that accompanied its presence, may have changed fertility behaviors throughout communities if the risk of Ebola appeared close to home. I hypothesize that the high risk for Ebola transmission during the epidemic in Sierra Leone led people to delay or postpone fertility by modifying social and reproductive behaviors, even if there was no direct biological risk to the individual.

I estimate the effect on aggregate district level fertility changes using difference-in-differences models and find a decrease of 0.4 births in districts 9 and 12 months after Ebola was first reported in the district. I also analyze a short-term reduction in fertility driven by individual timing of births using an event history analysis method. I find the birth intervals increased 16 percent for a woman residing in a district in which Ebola was reported compared to birth intervals for women in districts not exposed.

4.2 Background: Ebola Virus Disease and the 2014-2015 Outbreak in Sierra Leone

The first reported outbreak of Ebola was in 1976 in the Democratic Republic of Congo. Since then, there have been about a dozen outbreaks generally contained to less densely populated areas of central Africa (CDC 2017). The main symptoms of Ebola virus disease are headache, muscle pain, and fever followed by vomiting, diarrhea, and internal and external bleeding (WHO 2017). Ebola has high rates of direct person-to-person transmission via contact with body fluids (Gatherer 2014; Kucharski and Edmunds 2014). The zoonotic virus spreads rapidly between humans, causes severe hemorrhagic fever, and results in a case fatality rate of 60-70 percent (Kucharski and Edmunds 2014; Gatherer 2014). During the incubation period, Ebola-infected people are not contagious until they show symptoms, two to 21 days after infection (Chowell and Nishiura 2014).

The 2014-2015 West African Ebola epidemic started in Guinea in early 2014. In May of 2014, a nurse who attended a funeral in Guinea returned to her village in Kailahun district in Eastern Sierra Leone, where she infected six other nurses at the small clinic (World Health Organization 2014). Several of these women were transported to a hospital in Kenema, the largest city in the east of Sierra Leone.[72] As people travelled along major road networks between outlying eastern areas to cities such as the densely packed capital of Freetown in the west, Ebola rapidly spread to all fourteen districts across the country. By June 11, 2014, schools were closed and borders to Guinea and Liberia were shut to trade (Reuters 2014). On July 30, Sierra Leone President Ernest Bai Koroma issued a state of emergency for the country (Barbash 2014). The speed of the outbreak led to a high level of uncertainty and fear about the dangers, severity, and risks of Ebola.

The governments of Guinea, Sierra Leone, and Liberia and the World Health Organization encouraged people to avoid touching potentially-infected people and deceased people and to wash off bodily fluids with soap and water. Additional public health measures included curfews, restrictions on public meetings and gatherings, active surveillance and contract tracing for victims and potentially-infected persons, and a 30-day quarantine of potentially

[72] Information from personal field notes of interviews 2017 and from unpublished research conducted by Professor Claudena Skran, Lawrence University. See also (Richards et al. 2015; D. Schwartz, Abramowitz, and Anoko 2019).

infected individuals and their households (World Health Organization 2017). Regional lockdowns and travel restrictions were implanted on and off in response to numbers of cases, with nation-wide 3-day lockdowns in September 2014 and March 2015. These measures aimed to reduce the social contact people had during peaks of the outbreak. Negative and hostile social responses to these public health measures included removing loved ones from hospitals and conflicts between families performing burials and health workers (Richards et al 2015). Public health messaging was found to have promoted behavioral changes that supported efforts to contain the spread of the virus (Calain and Poncin 2015; Jalloh et al. 2020).

While individuals and families faced great economic uncertainty, Bowles et al (2016) find little association between declines in regional economic activities and Ebola cases in geographic areas outside of Monrovia (Liberia). Likewise, there were no sharp drops in formal unemployment in Sierra Leone, though it is unknown to what amount this extended to the informal employment sector where millions of Sierra Leoneans work (Casey, Glennerster, and Suri 2017). These studies find the strong associations at the national-level driven predominantly by international travel and trade restrictions (Games and Vickers 2015). The model I discuss below includes district fixed effects to capture regional economic, social, or political factors and time fixed effects to capture national changes over the course of the Ebola epidemic.

4.3 Literature: Infectious disease, risk perception, and fertility behavior modification

Demographic fertility theory establishes that reproductive behaviors and desires are complicated by epidemics and other external hardships (Boucekkine, Desbordes, and Latzer 2009). At a population level, fertility changes can either increase or decrease the total number of children born or shift the timing of births (Bongaarts and Feeney 1998). Long-term fertility changes align with large scale exogenous shocks such as social, political, or economic instability. Individuals may react to these shocks by delaying reproduction with the expectation that conditions favorable to family formation and childbearing will improve in the future, resulting in a baby bust (Caldwell 2004; Kohler and Kohler 2002; Sobotka, Skirbekk, and Philipov 2011; Blanc 2004; Agadjanian and Prata 2002; Lindstrom and Berhanu 1999). For example, armed conflicts have long been associated with changes in access to contraceptives and health care facilities, short and long-term displacement and marital separation, and economic uncertainty in ways that are associated with long-term fertility changes (McGinn et al. 2011; Thiede et al. 2020).

Short-term shocks can also change fertility behaviors, though such changes are not often sustained (Nobles, Frankenberg, and Thomas 2015; Lee Rodgers et al. 2005; Morin 2002; Scelfo 2002; Burlando 2014). Large-scale fertility changes result from the cumulative decisions of many individual couples. Therefore, causal associations between exogenous shocks (such as economic uncertainty and family separation) and behavior changes by couples (e.g. coital frequency) are difficult to disentangle.

Epidemics and other threats to health have been associated with fertility change due to biological and behavioral pathways. High rates of mortality across a population during a epidemic or pandemic can reduce the potential for future fertility (Carpenter et al. 1997; Underwood 1984); infectious diseases may also result in mortality at the fetal stage (Bloom-Feshbach et al. 2011). If there is a risk of disease exposure, medical professionals may encourage women planning to become pregnant to postpone childbearing due to increased risk or uncertainty of neonatal outcomes (Jamieson et al. 2006; Castro et al. 2018; Marteleto et al. 2017; Rangel, Nobles, and Hamoudi 2019).

Large demographic changes to fertility have been occurring in Africa in recent decades, and changes to fertility in Sierra Leone were already occurring before the 2014 Ebola outbreak started. The total fertility rate in Sierra Leone declined from over six children per women in 1960 to just over four children per women (World Bank 2017). Changes in fertility in Africa have not been adequately explained by Western and Eurocentric theories of fertility transition (Caldwell and Caldwell 1987; Bongaarts and Casterline 2012; Bledsoe, Banja, and Hill 1998). Culture, religion, and social norms affect the mechanisms through which information diffuses throughout a society, and fertility behaviors change as a result (Mason 1997; Bloom et al. 2008).

This theory of information diffusion also applies to the spread of public health information during the Ebola epidemic, indirectly affecting the fertility of women and couples through social and behavioral changes. The government-imposed lockdowns disrupted the norms of daily routine. Many young girls became pregnant while at home during the Ebola epidemic as schools were closed for nine months (Bandiera et al. 2019). The policies to reduce social interactions, such as closing schools and quarantine, may have changed the fertility behavior of young girls who would not necessarily have become pregnant without the shock of Ebola.

The perceived or actual threat of Ebola was found to have created a fear response in communities in Sierra Leone (Shultz et al. 2016). Information about the epidemic was predominantly communicated via radio. According to a qualitative study of journalists working in

Freetown, media coverage of Ebola served as a way to spread information about the epidemic and recommendations for staying safe and healthy during the uncertain time; over time, media became a collaborative space to create community unity and self-reflection in the face of disaster (Winters, Nordenstedt, and Alvesson 2020). The changing landscape of information availability, conflicting recommendations, and new discoveries made it difficult for citizens to feel confident or certain about what was fact, what was rumor, and what could keep them and their family safe.

One consequence is the under-utilization of health care facilities due to fear of health care staff, hospital conditions, or limited availability of supplies and personal. Additionally, disrupted access to reproductive health services and a lack of distribution of contraceptives decreased the ability to properly use modern contraception for family planning in Sierra Leone during the Ebola outbreak (Bietsch, Williamson, Reeves 2020). A report from 2018 about the Ebola outbreak in the Democratic Republic of the Congo suggests that couples were primarily using condoms to prevent pregnancy and also protect against sexual transmission of the Ebola virus; in the absence of condoms, some couples were temporarily resorting to natural contraception strategies such as abstinence and withdrawal to postpone pregnancy during the outbreak (McKay et al. 2019).

It is likely that information and fear about pregnancy risks due to diseases could be defused in similar patterns. Individual changes to fertility behavior must be widespread to have an impact on population level fertility. In general, during the epidemic, women were disproportionally at risk for contracting Ebola through caregiving activities (Menéndez et al. 2015; Diggins and Mills 2015). This increased the difficulty of accessing health care for infected and not-infected women. Pregnant women were acutely affected. Maternal mortality skyrocketed in West Africa during the Ebola outbreak due in part to the infection itself, but to a much greater extent due to the challenges on an already weak health care system in Sierra Leone (Vygen et al. 2016) . Already limited healthcare facilities and maternity units in rural areas were often closed due to a lack of staff and supplies needed to provide safe services to expecting mothers (Burkhardt, Erland, and Kahn 2019). Family planning services and supplies declined by 23 percent at the peak of the epidemic (Bietsch, Williamson, Reeves 2020). Documented again and again in the edited volume by Schwartz, Anoko, and Abramowitz (2019), stigmatization and fear led to hospitals and clinics turning away or delaying treatment for pregnant women who were unable to produce a certificate of a negative Ebola test. Pregnancy complications sometimes presented symptoms similar to Ebola. To expose an uninfected mother to other potentially infected patients was a risk to the mother and baby. But to unknowingly allow an infected mother into a clinic of uninfected pregnant women was a big risk for health care professionals. Many

preventable deaths of uninfected mothers were caused due to health centers being unable to cope with the high risk of Ebola exposure in clinics (Jones et al. 2016).

4.4 Data

4.4.1 Ebola Data

For data on the timing and intensity of Ebola, Backer and Wallinga (2016) compiled the WHO case reports published between December 2013 and June 2015, the peak of the epidemic. The weekly data counts the number of new reported cases in each district. I match this data to monthly fertility data (described below) based on the start date of May 25, 2014[73]. Sierra Leone was not determined to be clear of the epidemic until March 2016 because of the long treatment and recovery period for survivors.

The Ebola epidemic reached all corners of Sierra Leone, but the severity of impact varied drastically between districts. The cumulative number of cases also showed substantial variation between districts. There was heterogeneity in prevalence within districts; some villages were devastated while nearby villages remained completely case-free (Fang et al. 2016). One study in the districts Port Loko, Kambia, Moyamba and Pujehun suggests sub-district heterogeneity in local Ebola outbreaks was random (Bandiera et al. 2019); the spread of the virus clustered around new introductions of infection in patterns unrelated to urbanization, cultural, ethnic-linguistic, or climatological links (Gleason et al. 2017; Dudas et al. 2017). Figure 4.1 shows the variation in timing and severity of the Ebola by each of the 14 districts in Sierra Leone. The timing shows the spread of Ebola from the eastern Kailahun district, in the east where the first case was reported, through the middle of the country, to the western districts. The timing and severity were also driven by proximity to population centers such as Freetown in Western Area Urban district, Port Loko in Port Loko district, and Waterloo in Western Area Rural district.

4.4.2 Fertility Data

Data were obtained from the 2016 Sierra Leone Demographic and Health Malaria Indicator Survey (SL16 MIS) conducted by the National Malaria Control Program (2016). This is

[73] The data in the SL16 MIS is presented by month. If a week in the Backer and Wallinga data was a week that fell between two months, the month in the SL16 MIS was assigned based on which month had the most days of that particular week. For example, a week starting on January 31 was assigned to February because the majority of the days that week fell in February.

a household survey that is representative of married women ages 15-49 at the national, regional, and district levels. The SL16 MIS has a complete birth history of each woman for the five years prior to the survey, 2011-2016, which includes the 2014 Ebola outbreak. My analyses at the district level and at the individual level are weighted using the provided survey weights.

Descriptive Statistics of SL16 MIS

This analysis includes 8,501 women surveyed in the SL16 MIS; they reported a combined 6,890 births in the five years prior to the survey. All women in the sample were married. As Table 4.1 shows, motherhood is common among married woman in Sierra Leone: over 80% of the sampled women were mothers. Over half of the women in the sample reported having received no formal education. Universal primary education was instituted in 2004 (Government of Sierra Leone 2013). A greater proportion of younger women, specifically non-mothers, reported having been to school through secondary school. Women in the sample had an average of 3.2 children. Over half (57%) of the women in the sample reported giving birth in the five years prior to the survey (between 2011-2016). Of these births, half were male and 96% were singleton births. At the time of the survey, 96% of the children born were reported to be still living.

4.5 Method: Difference-in-Differences (DD)

The aggregate impact of the 2014 West African Ebola outbreak on fertility is tested by comparing districts where Ebola was present with districts that had no reported cases of Ebola at a given time, using a multi-period difference-in-differences (DD) model.

I examine the effect of Ebola on aggregate fertility using the district as the unit of analysis. As every district eventually reported cases of Ebola, treatment is rolled-out, resulting in a multi-period treatment. The DD models use the districts that do not yet have Ebola at a given time as the control or untreated category of comparison for the districts that do have Ebola at that same time. I have adjusted for differences in geography between the SL16 MIS and the Backer and Wallinga data[74]. The reports of Ebola in a district were operationalized as a binary variable where 1 is equal to at least one case of Ebola reported in a district in the 9 months and 12 months prior to a birth event (0 = otherwise). The operationalization of Ebola as a continuous variable

[74] For geographic consistency, the labels “Western Area Urban” and “Freetown” are the same.

yielded similar results as the main results; results are not reported as small counts in many districts at a given time presented challenges in interpretation.

There is heterogeneity across districts in the timing and number of cases present in a given month. Based on the number of cases, I create three treatment timing parameters: start, peak, and last. I identify the first week with reported cases of Ebola in a particular district; this is labeled as "start". The week with the maximum number of cases in that district is identified as the peak. The final time point is the last week for which any cases were reported. Estimations of fertility at these three time points test the hypothesis that the Ebola epidemic impacts fertility while accounting for differences across time. At the start of the epidemic, there was much more uncertainty and thus people may have modified their fertility behavior more than at other points.

Outcome: General Fertility Rate (GFR)

The births reported in the SL16 MIS are aggregated to the district level to produce a weighted count of the number of births in the district per month from 2011 to mid-2016. As the outcome of interest, the General Fertility Rate (GFR) is the aggregate measure of fertility (Equation 4.1). The monthly GFR is the number of births in a district *d*, in month *m*, divided by the total number of women of reproductive age (15-49) in a district. The denominator is constructed from the population-weighted number of married women surveyed in the SL16 MIS and assumed to be constant over time within the district (Equation 4.1). Weights are representative at the district level.

$$GFR = \frac{Weighted\ births\ in\ district\ in\ month}{Weighted\ women\ (15{-}49)\ in\ district} \qquad \text{(Equation 4.1)}$$

Timing: Fertility 9-month and 12-month lag

The month of each birth is reported in the SL16 MIS. I account for the time of conception by introducing a birth month lag of nine months, assuming no premature births. This would represent children conceived during the months of the Ebola outbreak and born nine months later. This conception lag is indicated as a "nine-month lag" throughout. I have also constructed a time variable that creates a "twelve-month lag" in births. This lag captures the period when a couple might be trying to or considering conceiving a child. I use the twelve-month lag to explore whether there were any earlier Ebola-related fertility modifications that may have influenced

fertility decisions due to repeated and constant information about the dangers of Ebola. Because the outbreak lasted for most of a year, some women may have been exposed to information about Ebola for several months before a potential conception. Additional analysis at eight, ten, eleven, and thirteen months were conducted to check robustness; I found similar trends, so results are not reported.

Parallel trends in GFR

The parallel trends assumption of DD requires parallel trends of fertility rates between districts prior to the Ebola epidemic. Figure 4.2 shows the patterns of the weighted general fertility rate in Sierra Leone by district. The red line represents the national average fertility rate during the same period. The blue bar on the X-axis represents the 2014 West African Ebola Outbreak and the green bar represents a nine-month conception delay to show the time that babies would be born if they were conceived during the months of the Ebola outbreak. There is an overall birth seasonality effect; the expected high fertility months in Sierra Leone are January through April (Dorélien 2016). Visually, all districts show trends that roughly align with the peaks and valleys of the seasonality as shown by the red line. Tests of parallel trends were conducted with sensitivity analysis of heterogeneity of treatment effects, and an event study specification (described below) shows non-significant differences in pre-treatment trends.

The Ebola outbreak eventually spread to every district in Sierra Leone. As a result, I will use a multi-period difference-in-difference estimator as the preferred specification:

$$GFR_{dm} = \alpha + \delta D_{dm} + \lambda_m + \eta_d + X + \varepsilon_{dm} \qquad \text{(Equation 4.2)}$$

GFR_{dm} is the outcome variable of the general fertility rate for a district d at month m. $\delta D_{d,m}$ is the parameter estimation δ for the dummy indicator D in district d at time m (in months). λ_m and η_d are the time- and district-fixed effects respectively. X includes controls for the calendar month to account for seasonality of birth in Sierra Leone[75].

[75] The month fixed effect captures trends over time (e.g. June 2013, July 2014, April 2015) while the seasonal fixed effect captures calendar months across all of the years of the analysis (e.g. June, July, April).

4.5.1 Difference in differences results

The DD models were run using both the twelve-month and nine-month fertility lag and at three different time points of Ebola cases in each district to account for differences in the first reported case of Ebola (start), the maximum number of cases (peak), and the final case (last). Results for the estimates of all six models are presented in Table 4.2. Each cell provides the δ estimation from a separate regression (robust standard errors in parentheses) with the model R^2 reported in italics. The sample size remained constant.

The interpretation of these results suggests that when using fixed effects for both district and time, the reports of the first cases of Ebola in a district led to a decrease the general fertility rate by 0.4 births in that district in the months following the start of the outbreak. In models adjusting for birth seasonality, this difference is not statistically significant at the 9-month lagged period for any time point. This suggests that the presence of the first reported Ebola case in a district does not have an immediate effect on conception rates in a given district. However, there is a significant decrease of almost half a birth on average in a district for births that occur 12 months after the first reported case of Ebola in a district[76]. This suggests that as the number of cases in a district increases, there is a decline in conceptions. The size of the fertility decline decreases as the Ebola epidemic continues; the effect on fertility is negative, although it is not statistically significant.

The average monthly GFR for all districts across the data was 7.39 weighted births per month (min 5.34, max 10.2). A decrease by 0.4 births is substantial. This result suggests a reduction of the fertility rate by almost half of a birth per month for women of reproductive age in a district after cases begin to increase in the district. Across all districts for all months Ebola was present, this results in approximately 92 fewer births for the mothers in the sample[77]. Given that there were 6,890 births in the SLMIS16 sample, up to 1% of all births could have been foregone because of Ebola. There was uncertainty and limited communication at the beginning of the Ebola epidemic that may have led to an atmosphere of fear and confusion. As the epidemic continued and more information was ascertained about the risk factors, treatment, and prevention, people may have returned to normal behavior if they perceived their risk to be low.

[76] The results from the preferred specification are included. Full analysis with and without fixed effects and calendar months are available upon request.
[77] Cumulatively across all districts there were 230 months where Ebola was present in a district.

Heterogeneity in Treatment Effects

The significant results for the nine-month lag and the twelve-month lag suggest that there may be heterogeneity in the treatment effect. Because of the long duration of the Ebola outbreak, behavior may have changed over time as the number of cases fluctuated. There also may have been differences based on the length of time Ebola was present in a district. Equation 4.3 puts the three time periods of Ebola (starting month to peak month (S), peak month to the last month (P), and the time period after the last month (L)) into a single DD model. The results of this model use the pre-Ebola period as the untreated period, and the periods after Ebola as the treated periods. The post-Ebola periods of peak-to-last and after-last are treated as two distinct time frames. Time and district fixed effects were included in this model. The model was run separately for the nine-month lag time and the twelve-month lag time.

$$GFR_{dm} = \alpha + \delta S_{dm} + \delta P_{dm} + \delta L_{dm} + \lambda_m + \eta_d + \varepsilon_{dm} \qquad \text{(Equation 4.3)}$$

Table 4.3 shows the results of the new DD models produced from Equation 4.3. These findings reinforce the interpretation that in the beginning of the Ebola outbreak in a district, the start-to-peak period, there was a decrease in fertility by almost half a birth. Fertility continued to decrease by half a birth during the period from the peak of the Ebola outbreak to the last case reported in a district. This suggests that throughout the entirety of the epidemic, people may have changed their fertility behaviors. Once a district was no longer reporting new cases of Ebola, there was not a lagged effect on fertility trends and possibly a return to normalcy.

An event study specification is an additional test of the robustness of the DD model. Using Equation 4.4, I test for heterogeneous treatment effects for the DD estimates with district $\eta_d$and month λ_m fixed effects, with *k* periods and *j* time intervals. Figure 4.3 shows the coefficients for the GFR for the three time points (start, peak, and last) at the nine-month and twelve-month lags in the twelve months prior to and after the start, peak, or end of Ebola. The vertical line at zero indicates the month Ebola started, peaked or ended in a particular district. The coefficients for measures of fertility compare each period to the zero-time point.

$$GFR_{dm} = \sum_{j=0}^{k2} \delta D_{dm} + \lambda_m + \eta_d + \varepsilon_{dm} \qquad \text{(Equation 4.4)}$$

There are no major patterns in the significance of the coefficients within any of the models[78]. However there appears to be a distinct pattern of declining fertility prior to the start of Ebola that was disrupted by the outbreak. Because the figure is centered on the start of Ebola for

[78] Results presented in Figure 4.3. Exact coefficients are available upon request.

each individual district, it is possible that the start of the Ebola outbreak in other districts influenced the start of the downward trend a few months prior to the start of Ebola in districts where the epidemic started later. However, even nationwide information of Ebola may not have induced people to change fertility behaviors until the outbreak was closer to home. The trends for the peak timing of Ebola suggest a pattern that is impacted by Ebola within a particular district's peak. Finally, the figures depicting fertility trends 9 and 12 months after the last case of Ebola show that the declines to fertility might have been temporary as the coefficients fluctuate around the zero after the end of Ebola, which is also influenced by data censure due to the timing of data collection.

The results from this difference-in-differences analysis suggest the Ebola outbreak had a negative impact on fertility rates in Sierra Leone. These results at the district level provide insight about how the outbreak affected fertility at different stages of the epidemic. The next analysis moves from examining the aggregate fertility changes to considering Ebola may have changed fertility on the individual.

4.6 Method: Event History Analysis

To examine the impacts of the West African Ebola outbreak on individual fertility, I use an event history analysis of the timing of births to show the decline in fertility found at the aggregate level in the difference-in-differences analysis may be produced by individual fertility decisions. This analysis uses the micro-level data of the SL16 MIS at the level of each birth in the birth history of each mother across time. Looking at the individual-level fertility decisions during the time of Ebola can provide a different understanding of how the outbreak was changing fertility behaviors of mothers. While the district-level analysis shows the degree of Ebola-related reduction in fertility, an examination of individual pregnancy trajectories allows a nuanced look at Ebola-related fertility decision-making. In the data, there is no indication of whether the mother was exposed to or infected with Ebola.

A Cox proportional hazard model (Equation 4.5) is used to compare births exposed to Ebola in the interval of conception (with a nine-month or twelve-month lag) to births not exposed to Ebola. The hazard ratio (HR) of births (i) exposed to Ebola (j) is adjusted for mother's characteristics such as age, educational level, the number of previous children born, and urban

status. The model also includes dummy variables for all of the districts in Sierra Leone and all of the month-year effects.

$$HR_{i,j} = e^{\beta_j} \quad \text{(Equation 4.5)}$$

To account for censuring data in the event history analysis, I limited my analysis to mothers who were age 15 or older. I also restricted my sample to fecund women who have ever had a first live birth (n=6,912 mothers). This limited my ability to look at younger girls and women who might be making a fertility decision for the first time, but it was a necessary precaution due to the lack of data about the timing of marriage or first intercourse in the survey. I analyzed the longitudinal data for every woman in the five-year time period between the start of the birth history and the end of the survey.

4.6.1 Event History Analysis Results

The relative risks of moving to a higher parity birth for mothers impacted by the Ebola outbreak during a 9- and 12-month lag in conception time is shown in Table 4.4. A birth is considered impacted if the interval between births includes months when Ebola was present in a district where the mother resides.

The results of the event history analysis report similar findings to the DD analysis. There was a 3.2% shorter interval between births born 9 months after Ebola was present in a district (RR= 1.032). This result is small and insignificant. However, there is a significantly greater relative risk at the 12-month lag interval (RR=0.842), suggesting a 16% longer birth interval in the 12 months after Ebola was present in a district. Thus, women were *less* likely to conceive another child during the time of the Ebola outbreak. This model excludes women who are having their first birth.

The mother's demographic characteristics controls mostly behave as expected. Older mothers have longer intervals between births and mothers with more children have shorter birth intervals (Gold et al. 2005). Increasing educational levels surprisingly shorten the interval between births in this analysis, but a small sample size for highly educated mothers may be driving this result, as the more-educated younger non-mothers are excluded from the analysis. Women in rural areas have 8.2% longer intervals between births if the interval is exposed to Ebola.

These results supplement the findings of the DD models by suggesting that the presence of Ebola in a district did not immediately produce changes in fertility behaviors. The effects are more pronounced 12 months after Ebola was first reported in a district, allowing for a 3-month period of uncertainty, and increasing cases, before a potential birth was conceived. Data collection in 2016 censures the analysis as some districts continued to report cases of Ebola into 2015. Data collection of more recent trends might show a diminishing effect as time goes on as the country and fertility behaviors either return to normal or adapt to a lower GFR and longer birth intervals.

4.7 Discussion

While this study cannot speak to specific biological effects of the spread of Ebola on fertility patterns, the findings of a population-level decrease in fertility and an individual level increase in the length of birth intervals suggest that reproductive behavior changes were responding to the spread of the virus. Measuring aggregate and individual changes to fertility during and after the 2014-2015 Ebola epidemic in Sierra Leone shows the longer-term effect of the epidemic on society. Longer birth intervals shed light on how social interactions and information spread can influence individual women's decisions that culminate as macro changes to district and national fertility patterns. Negative effects on fertility with the mere presence of Ebola in a district suggest that social interactions were disrupted during the epidemic regardless of a direct exposure or infection. These findings fit into the existing literature that people adjust reproductive behavior in times of uncertainty and disease spread.

This analysis is limited by three major factors: the SL16 MIS data about women, the geographic data available, and the timing of the data collection. First, the malaria indicator surveys (MIS) produced by the DHS are more limited than the standard DHS surveys. Thus, as mentioned before, the analysis is limited to married women and uses retrospective birth histories. It is possible that the results found in this analysis may overrepresent fertility changes as older mothers may be thinking about additional births differently than younger, unmarried women thinking about a first birth. A recent study of 5,775 young women ages 12-25 including both married and unmarried girls found that younger girls in villages in Sierra Leone highly disrupted by Ebola were more likely to spend more time with men and engage in risky, unprotected, or transactional sexual relationships (Bandiera et al. 2019). This evidence suggests that unmarried

women may face different pressures on fertility that may bias the direction of the results in opposite directions from older, married women. For example, school closures and increases in transactional relationships may increase fertility among younger, unmarried women while social isolation and decreased coital frequency may decrease fertility for older married women and mothers. Other potentially confounding variables such as health care utilization or contraceptive use are not reported in the MIS. The sample size also limits the ability to construct sub-group analysis similar to other papers (Grépin, Poirier, and Fox 2020). There is also no measure on desired fertility or Ebola exposure. As in any analysis of an infectious disease outbreak with a high fatality rate, there is a possibility of survival bias in retrospectively collected data.

The geographic unit used by the SL16 MIS and for reporting Ebola cases is the district. More detailed Ebola case counts at a smaller geographic unit exist (Fang et al. 2016). However, detailed retrospective birth and residency status information data is not available. It is unknown exactly to what extent international and internal migration contributed to the virus's spread. Movement across international boundaries for work and social visits challenged emergency public health efforts to contain the spread of the virus. Porous borders and seasonal labor migration, mostly among men, is common in West Africa (Roos 2014). The movement of people across boarders during the Ebola outbreak may have contributed to between 4 to 10 percent of new cases in adjacent districts (Backer and Wallinga 2016). Internal migration in Sierra Leone in 2014 and 2015 was limited by short-term state-wide lockdowns, economic pressures and temporary movements instead of being systematically associated with incidences of Ebola (World Bank Group 2015; Peak et al. 2018). Thus, in this analysis migration is assumed to be minimal in response to the nation-wide lock downs in place during Ebola. Other studies on female migration trends in Sierra Leone suggest that women tend to make short-term trips to visit family (World Bank Group 2015).

Similarly, data on rural or urban residence is available for women in the sample, but not differentiated in the Ebola data. Results in Dudas et al (2017) show that urban areas were not significantly associated with virus transmission; instead urban areas had higher case counts because of higher population figures. That is, the outbreak was more severe in urban areas because there were more reintroductions of the virus, but the spread of Ebola within a single introduction was similar in rural and urban areas. Therefore, the risks of coming in contact with someone in a city were higher because of population density, but the risks of contracting and possibly dying from Ebola were just as risky in rural areas as in urban areas once the virus was introduced. My finding that the interval between births for women in rural areas was longer than

for women in urban areas may suggest women in rural areas were responding to other factors such as decreased access to health care, infrequent contact with partners due to social isolation, or lower population density.

Finally, the collection of data so near to the end of the Ebola epidemic suggests that effects of delayed fertility may not yet be visible. The study design of censured data limits the ability to see births results a year after the end of the outbreak ended in particular districts. Additionally, the censured time cannot capture the long-term impact on the healthcare system and the long-term impact on Ebola survivors that Sierra Leone is still facing. It remains to be studied how the Ebola outbreak may change overall cohort sizes or family sizes in the long term.

The Ebola impact had small but significant negative effects on fertility in Sierra Leone. These results only scratch the surface of the short and long term effects of the Ebola outbreak on fertility patterns. New evidence shows negative effects on reproductive health among survivors including the potential for transmission of Ebola via vaginal fluid and breastmilk (Godwin et al. 2019). Medical knowledge gained during the West African outbreak, including new reports of the vertical transmission of the virus from mothers to unborn child, can strengthen understanding of the direct biological consequences of the virus on fertility for future outbreaks.

While the public health response from around the world was slow and criticized, Ebola was eventually contained and eliminated in West Africa (Delamou et al. 2017). However, since 2017 there have been five unrelated outbreaks of Ebola, primarily in the Democratic Republic of the Congo (Center for Disease Control and Prevention (CDC) 2019). These outbreaks have remained fairly isolated and a vaccination is being utilized, but the possibility of Ebola spreading and impacting larger populations is a potential reality. A vaccination may be a panacea, but currently there is little known on the effect of the vaccine on pregnant women (Gomes et al. 2017). Until then, mitigating the impact of Ebola and other viral disease will rely on public health campaigns for social and behavioral changes.

4.8 Tables

Table 4.1: Sample Size and descriptive statistics of the women age 15-49 in the Sierra Leone 2016 Demographic and Health Malaria Indicator Survey.

	All women	Mothers with a birth in last 5 years	Mothers without a birth in last 5 years	Non-mothers
Sample size (N)	8501	4887	3614	1589
Proportion of sample	1.00	0.57	0.43	0.19
Mean age	28.01	28.16	27.89	18.95
Education				
No Education	0.52	0.59	0.42	0.17
Primary	0.14	0.14	0.13	0.16
Secondary	0.33	0.26	0.43	0.66
Higher	0.01	0.01	0.02	0.02
Mean # of children ever born	3.37	3.86	2.71	-
Births in the last 5 years				
No births	0.43	-		
1 birth	0.37	0.65		
2 births	0.17	0.31		
3 births	0.02	0.04		
4 births	0.00	0.00		
Single Birth		0.96		
Male		0.50		
Still alive		0.96		

Source: SL16 MIS

Table 4.2: Impact of the 2014 West African Ebola outbreak on fertility in Sierra Leone Equation 4.1 for 9 and 12 months lagged models at three time points of Ebola in a district

Timing	**9 month lag models**	**12 month lag models**
Start	-0.405	-0.415*
	(-0.37)	(-0.22)
Peak	-0.009	-0.226*
	(-0.21)	(-0.12)
Last	-0.257	0.018
	(-0.16)	(-0.26)

The above table uses robust standard errors. Only the coefficients of interest are presented here. District and time fixed effect and calendar month coefficients are omitted. The sample size (n=932) remains constant across all estimates. Asterisks indicate * if $p \leq 0.1$, **if $p \leq 0.05$, ***if $p \leq 0.001$. R^2 reports are in grey italics.

Table 4.3: Impact of the 2014 Ebola Outbreak on fertility as a test of time varying treatment effects.

	9-Month Lag Model	12-Month Lag Model
Start to Peak (S)	-0.41	-0.40*
	(-0.37)	(-0.22)
Peak to Last (P)	-0.39	-0.61**
	(-0.46)	(-0.22)
After Last (L)	-0.643	-0.59
	(-0.46)	(-0.45)
Constant	0.93***	0.92***
	(-0.2)	(-0.2)
N	932	932
R^2	0.21	0.21

Note: Asterisks indicate * if $p \leq 0.1$, **if $p \leq 0.05$, ***if $p \leq 0.001$.

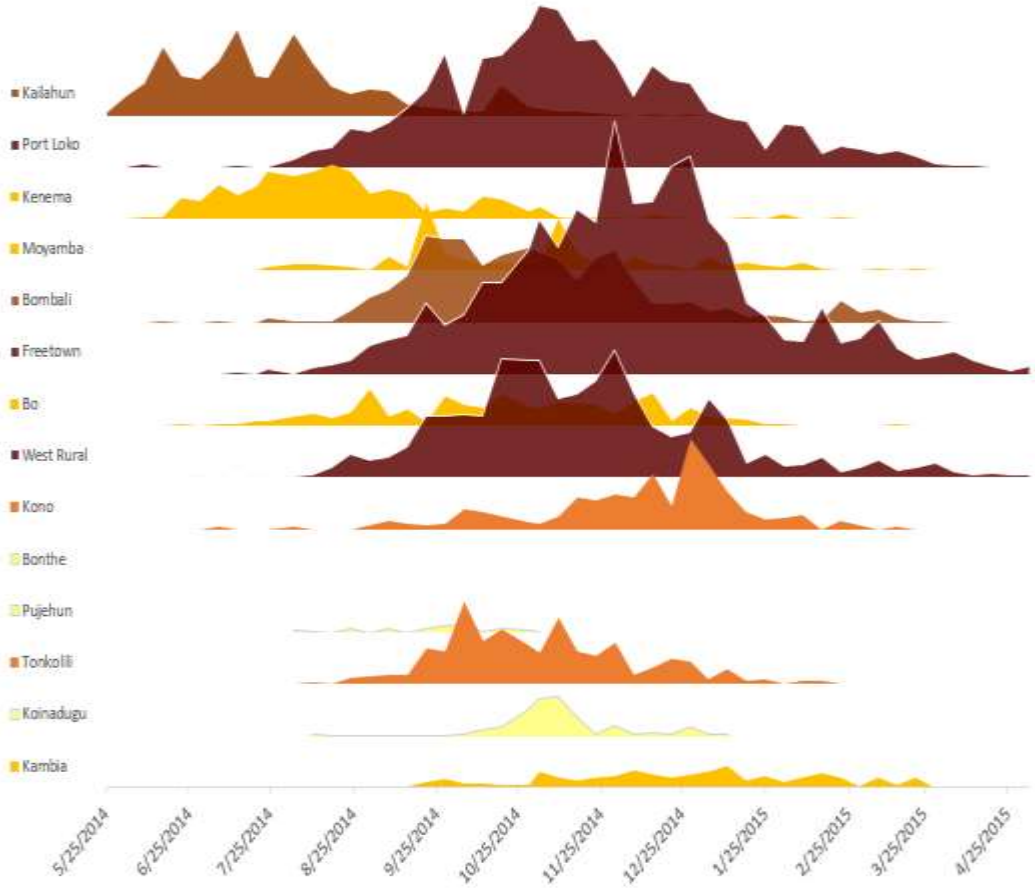
Table 4.4: Relative Risks of Increasing to Next Birth Parity

	9 Month Lag	12 Month lag
Ebola (1=yes)	1.032	
	(-0.079)	
Ebola (1=yes)		0.842*
		(-0.068)
Mother's age (ref. group= 15=19)		
20-24	0.727***	0.727***
	(-0.031)	(-0.031)
25-29	0.541***	0.541***
	(-0.026)	(-0.026)
30-34	0.422***	0.422***
	(-0.023)	(-0.023)
35-39	0.232***	0.232***
	(-0.016)	(-0.016)
40-45	0.117***	0.117***
	(-0.013)	(-0.013)
45-49	0.021***	0.021***
	(-0.007)	(-0.007)
Education (ref. group = none)		
Primary	1.041	1.041
	(-0.046)	(-0.046)
Junior Secondary	1.130**	1.130**
	(-0.051)	(-0.051)
Senior Secondary	1.164*	1.164*
	(-0.074)	(-0.074)
Technical	0.691*	0.691*
	(-0.116)	(-0.116)
Higher	1.088	1.088
	(-0.252)	(-0.252)
# of children	1.090***	1.090***
	(-0.007)	(-0.007)
Rural (ref. group = urban)	1.082+	1.083+
	(-0.044)	(-0.044)
N	479,740	479,740

Note: Asterisks indicate * if $p \leq 0.1$, **if $p \leq 0.05$, ***if $p \leq 0.001$.

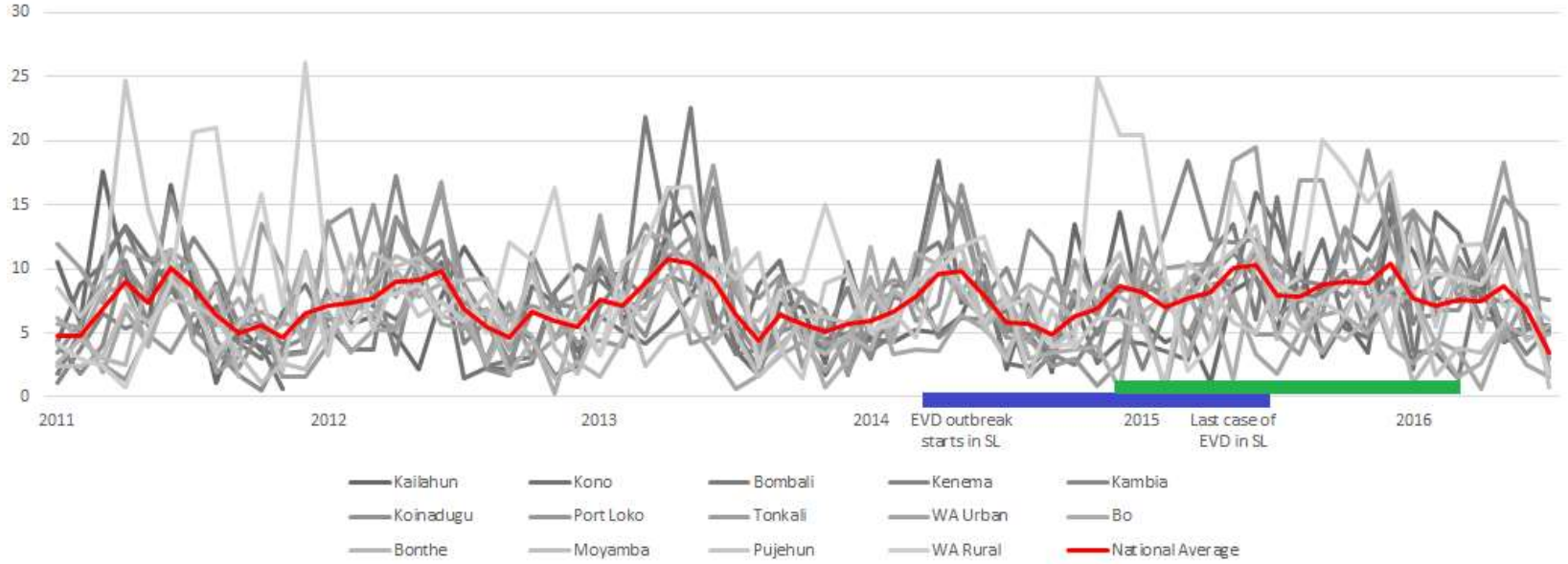
4.9 Figures

Figure 4.1: Number of cases reported in each district in Sierra Leone between May 2014, and April 2015.



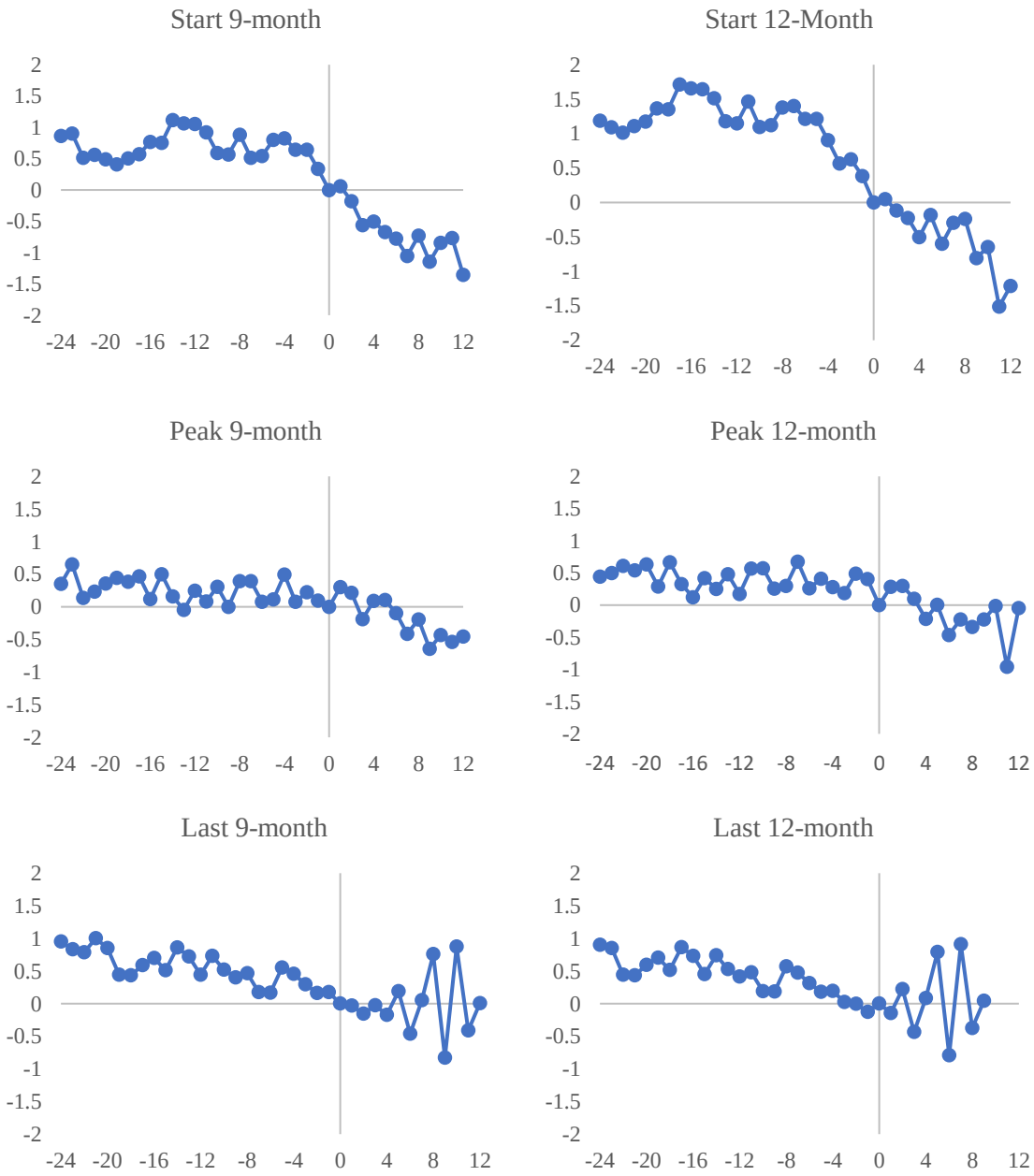
Source: Author's calculations using data from Backer and Wallinga, (2016)

Figure 4.2: Birth Trends in Sierra Leone (GFR) from the SL16 MIS



Legend: The blue bar on the X-axis represents the duration of the 2014 West African Ebola Outbreak. The green bar represents the duration of the Ebola outbreak offset by nine months to represent the delay between conception and birth.

Figure 4.3: Heterogeneity of Effects of Ebola Outbreak on Fertility Over Time



Coefficients on y-axis. Vertical line at 0 represents the time point specific to the graph (Start, Peak, or Last). X-axis represents months prior to and after the specified time point.

References

Abelsæth, Anne. 2012. “Tutorial: Development of Data Entry- and CAPI Applications in CSPro.” Notater Documents. Statistics Norway.

Agadjanian, Victor, and Ndola Prata. 2002. “War, Peace, and Fertility in Angola.” *Demography* 39 (2): 215–31.

Alam, Shamma Adeeb. 2015. “Parental Health Shocks, Child Labor and Educational Outcomes: Evidence from Tanzania.” *Journal of Health Economics* 44: 161–75.

Aldersey, Heather M. 2012. “Family Perceptions of Intellectual Disability: Understanding and Support in Dar Es Salaam.” *African Journal of Disability* 1 (1): 1–12.

Aldersey, Heather M., and H. Rutherford Turnbull. 2011. “The United Republic of Tanzania’s National Policy on Disability: A Policy Analysis.” *Journal of Disability Policy Studies* 22 (3): 160–69.

Altman, Barbara M., and Sharaon N. Barnartt, eds. 2006. International Views on Disability Measures: Moving Toward Comparative Measurement. Elsevier LtD.

American Association for Public Opinion Research. 2016. “Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th Edition.”

Avieli, Hila, Tova Band-Winterstein, and Tal Araten Bergman. 2019. “Sibling Relationships Over the Life Course: Growing Up With a Disability.” *Qualitative Health Research* 29 (12): 1739–50.

Backer, Jantien A., and Jacco Wallinga. 2016. “Supplemental Information: Spatiotemporal Analysis of the 2014 Ebola Epidemic in West Africa.” *PLoS Computational Biology* 12 (12): 1–19.

Bakhshi, Parul, Ganesh M. Babulal, and Jean Francois Trani. 2020. “Disability, Poverty, and Schooling in Post-Civil War in Sierra Leone.” *European Journal of Development Research*.

Bandiera, Oriana, Niklas Buehren, Markus Goldstein, Imran Rasul, and Andrea Smurra. 2019. “The Economic Lives of Young Women in the Time of Ebola: Lessons from an Empowerment Program.” Report December.

Banerjee, Abhijit V., and Esther Duflo. 2011. “Top of the Class.” In *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: PublicAffairs.

Bannink, Femke, Ruth Nalugya, and Geert van Hove. 2020. “‘They Give Him a Chance’-Parents’ Perspectives on Disability and Inclusive Primary Education in Uganda.” *International Journal of Disability, Development and Education* 67 (4): 357–75.

Barbash, Fred. 2014. “Sierra Leone Declares State of Emergency in Ebola Crisis.” The Washington Post. http://www.washingtonpost.com/news/morning-mix/wp/2014/07/31/sierra-leone-declares-state-of-emergency-in-ebola-crisis/.

Barrera-Osorio, Felipe, Leigh L. Linden, and Francisco Perez-Calle. 2008. “Conditional Cash Transfers in Education Design Features, Peer and Sibling Effects: Evidence from a Randomized Experiment in Colombia.” NBER Working Paper Series.

Bauer, Johannes J. 2016. “Biases in Random Route Surveys.” *Journal of Survey Statistics and Methodology* 4 (2): 263–87.

Beerten, Roeland, Peter Lynn, Johanna Laiho, and Jean Martin. 2014. “Response Rates as a Measure of Survey Quality.”

Begum, Gazi, and Jan Blacher. 2011. “The Sibling Relationship of Adolescents with and without Intellectual Disabilities.” *Research on Developmental Disabilities* 32(5): 1580–88.

Bennett, Carol, Sara Khangura, Jamie C. Brehaut, Ian D. Graham, David Moher, Beth K. Potter, and Jeremy Grimshaw. 2011. “Reporting Guidelines for Survey Research: An Analysis of Published Guidance and Reporting Practices.” *PLoS Medicine* 8 (8): 1–11.

Bennett, Steve, Tony Woods, Winitha M. Iiyanagec, and Duane L. Smith. 1991. “A Simplified General Method for Cluster-Sample Surveys of Health in Developing Countries.” *World Health Stat Q*. 44(3): 98–106.

Berman, Elizabeth Popp, and Daniel Hirschman. 2018. “The Sociology of Quantification: Where Are We Now?” *Contemporary Sociology: A Journal of Reviews* 47(3): 257–66.

Bietsch, Kristin, Jessica Williamson, and Margaret Reeves. 2020. “Family Planning During and After the West African Ebola Crisis.” *Studies in Family Planning* 51 (1): 71–86. https://doi.org/10.1111/sifp.12110.

Bischoff, Lisa G, and Daniel H Tingstrom. 1991. “Siblings of Children with Disabilities: Psychological and Behavioural Characteristics.” *Counselling Psychology Quarterly* 4 (4): 311–21.

Black, Sandra E, Sanni Breining, David N Figlio, Jonathan Guryan, Krzysztof Karbownik, Helena Skyt Nielsen, Jeffrey Roth, and Marianne Simonsen. 2020. “Sibling Spillovers*.” *The Economic Journal*, no. 2012: 1–28.

Blanc, A K. 2004. “The Role of Conflict in the Rapid Fertility in Eritrea and Prospects for the Future.” *Studies in Family Planning* 35 (4): 236–45.

Bledsoe, Caroline, Fatoumatta Banja, and Allan G. Hill. 1998. “Reproductive Mishaps and Western Contraception: An African Challenge to Fertility Theory.” *Population & Development Review* 24 (1): 15–57.

Blom, Annelies G. 2009. “Nonresponse Bias Adjustments: What Can Process Data Contribute?” Social Sciences. ISER Working Paper Series

Bloom-Feshbach, Kimberly, Lone Simonsen, Cécile Viboud, Kåre Mølbak, Mark A. Miller, Magnus Gottfredsson, and Viggo Andreasen. 2011. “Natality Decline and Miscarriages Associated with the 1918 Influenza Pandemic: The Scandinavian and United States Experiences.” *Journal of Infectious Diseases* 204 (8): 1157–64.

Bold, Tessa, Deon Filmer, Gayle Martin, Brian Stacy, Christophe Rockmore, Jakob Svensson, and Waly Wane. 2017. “Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa.” *Journal of Economic Perspectives* 31 (4): 185–204.

Bongaarts, John, and Griffith Feeney. 1998. “On the Quantum and Tempo of Fertility.” *Population and Development Review* 24 (2): 271.

Bongaarts, John, and John Casterline. 2012. “Fertility Transition: Is Sub-Saharan Africa Different? Fertility Levels and Trends.” *Population & Development Review* 38: 153–68.

Borgers, Natacha, Edith De Leeuw, and Joop Hox. 2000. “Children as Respondents in Survey Research: Cognitive Development and Response Quality.” *Bulletin de Methodolgie Sociologique* (66): 60–75.

Boucekkine, Raouf, Rodolphe Desbordes, and Hélène Latzer. 2009. “How Do Epidemics Induce Behavioral Changes?” *Journal of Economic Growth* 14 (3): 233–64.

Bowles, Jeremy, Jonas Hjort, Timothy Melvin, and Eric Werker. 2016. “Ebola, Jobs and Economic Activity in Liberia.” *Journal of Epidemiology and Community Health* 70(3): 271–77.

Brackertz, Nicola. 2007. “Who Is Hard to Reach and Why?” ISR Working Paper Series.

Bronfenbrenner, Urie, and P.A. Morris. 2006. “The Bioecological Model of Human Development.” In *Handbook of Child Psychology, Vol. 1: Theoretical Models of Human Development*, edited by W. Damon and R. M. Lerner, 6th ed., 793–828. New York: Wiley.

Bronfenbrenner, Urie. 1986. “Ecology of the Family as a Context for Human Development.” *Developmental Psychology* 22 (6): 723–42.

Burke, Meghan M., Catherine K. Arnold, and Aleksa L. Owen. 2015. “Sibling Advocacy: Perspectives About Advocacy from Siblings of Individuals with Intellectual and Developmental Disabilities.” *Inclusion* 3 (3): 162–75.

Burkhardt, Gillian, Elin Erland, and Patricia Kahn. 2019. “Nowhere to Go: The Challenges of Caring for Pregnant Women in Freetown During Sierra Leone’s Ebola Virus Epidemic.” In *Pregnant in the Time of Ebola: Women and Their Children in the 2013-2015 West African Epidemic*, edited by David A. Schwartz, Julienne Ngoundoung Anoko, and Sharon A. Abramowitz, 295–312. Cham, Switzerland: Springer Nature Switzerland AG.

Burlando, Alfredo. 2014. “Power Outages, Power Externalities, and Baby Booms.” *Demography* 51 (4): 1477–1500.

Calain, Philippe, and Marc Poncin. 2015. “Reaching out to Ebola Victims: Coercion, Persuasion or an Appeal for Self-Sacrifice?” *Social Science and Medicine* 147: 126–33.

Caldwell, John C, and Pat Caldwell. 1987. “The Cultural Context of High Fertility in Sub-Saharan Africa.” *Population and Development Review* 13 (3): 409–37.

Caldwell, John C. 2004. “Social Upheaval and Fertility Decline.” *Journal of Family History* 29 (4): 382–406.

Carley-Baxter, Lisa R, Craig A Hill, David J Roe, Susan E Twiddy, Rodney K Baxter, and Jill Ruppenkamp. 2009. “Does Response Rate Matter? Journal Editors Use of Survey Quality Measures in Manuscript Publication Decisions.” *Survey Practice* 2 (7).

Carpenter, Lucy M., Jessica S. Nakiyingi, Anthony Ruberantwari, Samuel S. Malamba, Anatoli Kamali, and James a G Whitworth. 1997. “Estimates of the Impact of HIV Infection on Fertility in a Rural Ugandan Population Cohort.” *Health Transition Review* 7 (2): 113–26.

Casey, Katherine, Rachel Glennerster, and Tavneet Suri. 2017. “The Economic Impacts of Ebola on Firms in Sierra Leone.” *International Growth Centre* 5 (12): 542–50. http://blogs.worldbank.org/education/impact-ebola-education-sierra-leone.

Castro, Marcia C., Qiuyi C. Han, Lucas R. Carvalho, Cesar G. Victora, and Giovanny V. A. França. 2018. “Implications of Zika Virus and Congenital Zika Syndrome for the Number of Live Births in Brazil.” *Proceedings of the National Academy of Sciences* 115 (24): 6177–82.

Caviglia-Harris, Jill, Simon Hall, Katrina Mulllan, Charlie Macintyre, Simone Carolina Bauch, Daniel Harris, Erin Sills, Dar Roberts, Michael Toomey, and Hoon Cha. 2012. “Improving Household Surveys through Computer-Assisted Data Collection: Use of Touch-Screen Laptops in Challenging Environments.” *Field Methods* 24 (1): 74–94.

Center for Disease Control and Prevention (CDC). 2019. “Years of Ebola Virus Disease Outbreaks.” *U.S. Department of Health & Human Services*. 2019.

Chamie, Mary. 1989. “Survey Design Strategies for the Study of Disability.” *World Health Statistics Quarterly*. 42 (3): 122–40.

Chataika, Tsitsi, and Judy Mckenzie. 2013. “Childhood Disability Studies.” Disabled Children’ *Childhood Studies*, 152–53.

Chen, Xinguang, Hui Hu, Xiaohui Xu, Jie Gong, Yaqiong Yan, and Fang Li. 2018. “Probability Sampling by Connecting Space with Households Using GIS/GPS Technologies.” *Journal of Survey Statistics and Methodology* 6 (2): 149–68.

Chitiyo, Morgan, and George Chitiyo. 2007. “Special Education in Southern Africa: Current Challenges and Future Threats.” *The Journal of the International Association of Special Education* 8 (1): 61–68.

Chowell, Gerardo, and Hiroshi Nishiura. 2014. “Transmission Dynamics and Control of Ebola Virus Disease: A Review.” *BMC Medicine* 12: 196.

Cicirelli, Victor G. 1994. “Sibling Relationships in Cross-Cultural Perspective.” *Journal of Marriage and Family* 56 (1): 7–20.

Couper, Mick P. 2005. “Technology Trends in Survey Data Collection.” *Social Science Computer Review* 23 (4): 486–501.

Cox, Martha J, and Blair Paley. 2003. “Understanding Families as Systems.” *Current Directions in Psychological Science* 12 (5): 193–96.

Croft, Alison. 2013. “Promoting Access to Education for Disabled Children in Low-Income Countries: Do We Need to Know How Many Disabled Children There Are?” *International Journal of Educational Development* 33 (3): 233–43.

Crowne, Douglas P., and David A. Marlowe. 1960. “A New Scale of Social Desirability Independent of Pathology.” *Journal of Consulting Psychology* 24: 351.

Deen, Gibrilla F., Barbara Knust, Nathalie Broutet, Foday R. Sesay, Pierre Formenty, Christine Ross, Anna E. Thorson, et al. 2015. “Ebola RNA Persistence in Semen of Ebola Virus Disease Survivors — Preliminary Report.” *New England Journal of Medicine*, 151014140118009.

Delamou, Alexandre, Thérèse Delvaux, Alison Marie El Ayadi, Abdoul Habib Beavogui, Junko Okumura, Wim Van Damme, and Vincent De Brouwere. 2017. "Public Health Impact of the 2014–2015 Ebola Outbreak in West Africa: Seizing Opportunities for the Future." *BMJ Global Health* 2 (2): e000202.

Dew, Angela, Susan Balandin, and Gwynnyth Llewellyn. 2008. "The Psychosocial Impact on Siblings of People with Lifelong Physical Disability: A Review of the Literature." *Journal of Developmental Physical Disabilities* 20: 485–507.

DHS Program. 2021. "DHS Methodology." The DHS Program: Demographic and Health Surveys. 2021. https://dhsprogram.com/Methodology/Survey-Types/DHS-Methodology.cfm.

Diggins, Jennifer, and Elizabeth Mills. 2015. "The Pathology of Inequality: Gender and Ebola in West Africa." IDS Practice Paper in Brief 23 February (February).

Dooblo. "SurveyToGo." https://www.dooblo.net/.

Doody, Mairéad A, Richard P Hastings, Sarah O Neill, and Ian M Grey. 2010. "Research in Developmental Disabilities Sibling Relationships in Adults Who Have Siblings with or without Intellectual Disabilities" *Research in Developmental Disabilities* 31: 224–31.

Dorélien, Audrey M. 2016. "Birth Seasonality in Sub-Saharan Africa." *Demographic Research* 34 (1): 761–96.

Dudas, Gytis, Luiz Max Carvalho, Trevor Bedford, Andrew J. Tatem, Guy Baele, Nuno R. Faria, Daniel J. Park, et al. 2017. "Virus Genomes Reveal Factors That Spread and Sustained the Ebola Epidemic." *Nature* 544 (7650): 309–15.

Elston, J. W.T., C. Cartwright, P. Ndumbi, and J. Wright. 2017. "The Health Impact of the 2014–15 Ebola Outbreak." *Public Health* 143: 60–70.

Escamilla, Veronica, Michael Emch, Leonard Dandalo, William C Miller, and Irving Hoffman. 2014. "Sampling at Community Level by Using Satellite Imagery and Geographical Analysis." *Bulletin of the World Health Organization* 92 (June): 690–94.

Espeland, Wendy Nelso, and Mitchell L. Stevens. 2008. "A Sociology of Quantification." *European Journal of Sociology*: 401–36.

Esri. 2019. "World Street Map." https://www.arcgis.com/home/item.html?id=3b93337983e9436f8db950e38a8629af.

Evans, Ruth. 2010. "Children's Caring Roles and Responsibilities within the Family in Africa." *Geography Compass* 4 (10): 1477–96.

Fairhead, James. 2014. "The Significance of Death, Funerals, and the after-Life in Ebola-Hit Sierra Leone, Guinea and Liberia: Anthropological Insights into Infection and Social Resistance," Working Paper. October: 5. http://www.heart-resources.org/wp-content/uploads/2014/10/FairheadEbolaFunerals8Oct.pdf.

Fakhoury, Walid, and Stefan Priebe. 2002. "The Process of Deinstitutionalization: An International Overview." *Current Opinions in Psychiatry* 15: 187–92.

Fallah, Mosoka P., Laura A. Skrip, Bernice T. Dahn, Tolbert G. Nyenswah, Hilary Flumo, Meekie Glayweon, Tee L. Lorseh, Stephen G. Kaler, Elizabeth S. Higgs, and Alison P.

Galvani. 2016. “Pregnancy Outcomes in Liberian Women Who Conceived after Recovery from Ebola Virus Disease.” *The Lancet Global Health* 4 (10): e678–79.

Fang, Li Qun, Yang Yang, Jia Fu Jiang, Hong Wu Yao, David Kargbo, Xin Lou Li, Bao Gui Jiang, et al. 2016. “Transmission Dynamics of Ebola Virus Disease and Intervention Effectiveness in Sierra Leone.” *Proceedings of the National Academy of Sciences of the United States of America* 113 (16): 4488–93.

Farmer, Paul. 2020. Fever, Feuds, and Diamonds: Ebola and the Ravages of History. Farrar, Straus and Giroux.

Filmer, Deon, and Kinnon Scott. 2012. “Assessing Asset Indices.” *Demography* 49 (1): 359–92. https://doi.org/10.1007/s13524-011-0077-5.

Filmer, Deon. 2008. “Disability, Poverty, and Schooling in Developing Countries: Results from 14 Household Surveys.” *World Bank Economic Review* 22 (1): 141–63.

Fincham, Jack E. 2008. “Response Rates and Responsiveness for Surveys, Standards, and the Journal.” *American Journal of Pharmaceutical Education* 72 (2): 43.

Fitzgerald, Robert, and Linda Fuller. 1982. “I Hear You Knocking But You Can’t Come In: The Effect of Reluctant Respondents and Refusers on Sample Survey Estimates.” *Sociological Methods & Research* 11(1): 3–32.

Fletcher, Jason, Nicole L Hair, and Barbara L Wolfe. 2012. “Am I My Brother’s Keeper? Sibling Spillover Effects: The Case of Developmental Disabilities and Externalizing Behavior.” National Bureau of Economic Research Working Paper Series No. 18279.

Fotso, Arlette Simo. 2017a. “Child Disability and Mothers’ Labour Market Participation in Cameroon.” *Journal of African Development* 19 (1): 27–60.

Fotso. 2017b. “Child Disability and Siblings’ Healthcare Expenditures in a Context of Child Fostering.” *Social Science and Medicine* 182: 89–96.

Fottrell, Edward, and Peter Byass. 2008. “Population Survey Sampling Methods in a Rural African Setting: Measuring Mortality.” *Population Health Metrics* 6. https://doi.org/10.1186/1478-7954-6-2.

Fujiura, Glenn T., and Kiyoshi Yamaki. 2000. “Trends in Demography of Childhood Poverty and Disability.” *Exceptional Children* 66 (2): 187–99.

Fujiura, Glenn T., Hye J. Park, and Violet Rutkowski-Kmitta. 2005. “Disability Statistics in the Developing World: A Reflection on the Meanings in Our Numbers.” *Journal of Applied Research in Intellectual Disabilities* 18(4): 295–304.

Fujiura, Glenn T., Violet Rutkowski-Kmitta, and Randall Owen. 2010. “Make Measurable What Is Not so: National Monitoring of the Status of Persons with Intellectual Disability.” *Journal of Intellectual and Developmental Disability* 35(4): 244–58.

Galdo, Jose, Ana C Dammert, and Degnet Abebaw. 2019. “Child Labor Measurement in Agricultural Households: Seasonality, Proxy Respondent and Gender Information Gaps in Ethiopia.” Bonn, Germany. https://glm-lic.iza.org/wp-content/uploads/2018/06/glmlic-wp043.pdf.

Galea, Sandro, and Melissa Tracy. 2007. “Participation Rates in Epidemiologic Studies.” *Annals of Epidemiology* 17 (9): 643–53.

Galway, Lindsay P., Nathaniel Bell, Sahar A.E. Al Shatari, Amy Hagopian, Gilbert Burnham, Abraham Flaxman, Wiliam M. Weiss, Julie Rajaratnam, and Tim K. Takaro. 2012. "A Two-Stage Cluster Sampling Method Using Gridded Population Data, a GIS, and Google Earth TM Imagery in a Population-Based Mortality Survey in Iraq." *International Journal of Health Geographics* 11: 1–9.

Games, Dianna, and Brendan Vickers. 2015. "The Ebola Crisis in West Africa: Implications for Trade and Regional Integration." Commonwealth Trade Hot Topics 119.

Gatherer, Derek. 2014. "The 2014 Ebola Virus Disease Outbreak in West Africa." *Journal of General Virology* 95 (PART 8): 1619–24.

Gelman, Andrew. 2007. "Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22 (2): 153–64.

Gidigo, S, AM Olademeji, AA Roberts, P Nguku, IG Nwangwu, NE, Shuaib F, et al. 2015. "Public Knowledge, Perception and Source of Information on Ebola Virus Disease." *PLOS Currents Outbreaks*, 1–12.

Gleason, B. L., S. Foster, G. E. Wilt, B. Miles, B. Lewis, K. Cauthen, M. King, et al. 2017. "Geospatial Analysis of Household Spread of Ebola Virus in a Quarantined Village - Sierra Leone, 2014." *Epidemiology and Infection* 145 (14): 2921–29.

Godwin, Christine L., David A. Wohl, William A. Fischer, Kavita Singh, Darrell A. Hawks, Elizabeth E. Devore, and Jerry Brown. 2019. "Reproductive Health Sequelae among Women Who Survived Ebola Virus Disease in Liberia." *International Journal of Gynecology and Obstetrics* 146 (2): 212–17.

Gold, Rachel, Frederick A. Connell, Patrick Heagerty, Peter Cummings, Stephen Bezruchka, Robert Davis, and Mary Lawrence Cawthon. 2005. "Predicting Time to Subsequent Pregnancy." *Maternal and Child Health Journal* 9 (3): 219–28.

Gomes, Melba F., Vânia De La Fuente-Núñez, Abha Saxena, and Annette C. Kuesel. 2017. "Protected to Death: Systematic Exclusion of Pregnant Women from Ebola Virus Disease Trials." *Reproductive Health* 14 (Suppl 3). https://doi.org/10.1186/s12978-017-0430-2.

Gosen, Steganie. 2014. "Social Desirability in Survey Research: Can the List Experiment Provide the Truth?" Dissertation for a Doctorate of Natural Science from University of Marburg, Germany.

Graff, J. Carolyn, Susan Neely-Barnes, and Heather Smith. 2008. "Theoretical and Methodological Issues in Sibling Research." In *International Review of Research in Mental Retardation*, edited by Laraine Masters Glidden, 36:233–80. Elsevier Inc.

Grahe, Jon. 2018. "Another Step towards Scientific Transparency: Requiring Research Materials for Publication." *Journal of Social Psychology* 158 (1): 1–6.

Grais, Rebecca F, Angela M C Rose, and Jean-Paul Guthmann. 2007. "Don't Spin the Pen: Two Alternative Methods for Second-Stage Sampling in Urban Cluster Surveys." *Emerging Themes in Epidemiology* 4 (8). https://doi.org/10.1186/1742-7622-4-8.

Graner, Elvira. 2001. "Labor Markets and Migration in Nepal." *Mountain Research and Development* 21 (3): 253–59.

Grech, Shaun. 2015. "Decolonising Eurocentric Disability Studies: Why Colonialism Matters in the Disability and Global South Debate." *Social Identities* 21 (1): 6–21.

Green, Sara, Christine Davis, Elana Karshmer, Pete Marsh, and Benjamin Straight. 2005. “Living Stigma: The Impact of Labeling, Stereotyping, Separation, Status Loss, and Discrimination in the Lives of Individuals with Disabilities and Their Families.” *Sociological Inquiry* 75 (2): 197–215.

Grépin, Karen A., Mathieu J.P. Poirier, and Ashley M. Fox. 2020. “The Socio-Economic Distribution of Exposure to Ebola: Survey Evidence from Liberia and Sierra Leone.” SSM - Population Health 10.

Groce, Nora Ellen. 2003. “Adolescents and Youth with Disability: Issues and Challenges.” Working Paper.

Grover, Sonja. 2004. “Why Won’t They Listen to Us? On Giving Power and Voice to Children Participating in Social Research.” *Childhood* 11(1): 81–93.

Groves, Robert M. 2011. “Three Eras of Survey Research.” *Public Opinion Quarterly* 75 (5 SPEC. ISSUE): 861–71.

Groves, Robert M., and Emilia Peytcheva. 2008. “The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis.” *Public Opinion Quarterly* 72 (2): 167–89.

Groves, Robert M., and Lars Lyberg. 2010. “Total Survey Error: Past, Present, and Future.” *Public Opinion Quarterly* 74 (5): 849–79.

Groves, Robert M., Robert B. Cialdini, and Mick P. Couper. 1992. “Understanding the Decision to Participate in a Survey.” *Public Opinion Quarterly* 56: 475–95.

Haenssgen, Marco J. 2015. “Satellite-Aided Survey Sampling and Implementation in Low- and Middle-Income Contexts: A Low-Cost/Low-Tech Alternative.” *Emerging Themes in Epidemiology* 12 (1): 1–10.

Hall, Sarah A., and Zach Rossetti. 2018. “The Roles of Adult Siblings in the Lives of People with Severe Intellectual and Developmental Disabilities.” *Journal of Applied Research in Intellectual Disabilities* 31 (3): 423–34.

Hansen, M., M. Harty, and J. Bornman. 2016. “Exploring Sibling Attitudes towards Participation When the Younger Sibling Has a Severe Speech and Language Disability.” *South African Journal of Child Health* 10 (1): 47–51.

Harling, Guy, Dumile Gumede, Tinofa Mutevedzi, Nuala McGrath, Janet Seeley, Deenan Pillay, Till W. Bärnighausen, and Abraham J. Herbst. 2017. “The Impact of Self-Interviews on Response Patterns for Sensitive Topics: A Randomized Trial of Electronic Delivery Methods for a Sexual Behaviour Questionnaire in Rural South Africa.” *BMC Medical Research Methodology* 17 (1): 1–14.

Harry, Beth. 2002. “Trends and Issues in Serving Culturally Diverse Families of Children with Disabilities.” *The Journal of Special Education* 36 (3): 132–40.

Hartley, Sally, P. Ojwang, A. Baguwemu, M. Ddamulira, and A. Chavuta. 2005. “How Do Carers of Disabled Children Cope? The Ugandan Perspective.” *Child Care, Health and Development* 31 (2): 167–80.

Hartling, Lisa, Andrea Milne, Lisa Tjosvold, Dawn Wrightson, Jennifer Gallivan, and Amanda S. Newton. 2014. “A Systematic Review of Interventions to Support Siblings of Children with Chronic Illness or Disability.” *Journal of Pediatrics and Child Health* 50 (10): E26–38.

Hastings, R. P. 2016. Do Children With Intellectual and Developmental Disabilities Have a Negative Impact on Other Family Members? The Case for Rejecting a Negative Narrative. *International Review of Research in Developmental Disabilities*. Vol. 50. Elsevier Ltd.

Hedges, Sophie, Rebecca Sear, Jim Todd, Mark Urassa, and David Lawson. 2019. "Earning Their Keep? Fostering, Children's Education, and Work in North-Western Tanzania." *Demographic Research* 41 (July): 263–92.

Helmes, Edward, and Ronald R Holden. 2003. "The Construct of Social Desirability: One or Two Dimensions?" *Personality and Individual Differences* 34: 1015–23.

Henderson, Ralph H., Hillard Davis, Donald L. Eddins, and William H. Foege. 1973. "Assessment of Vaccination Coverage, Vaccination Scar Rates, and Smallpox Scarring in Five Areas of West Africa." Bulletin of the World Health Organization 48 (2): 183–94.

Hill, Malcolm, John Davis, Alan Prout, and Kay Tisdall. 2004. "Moving the Participation Agenda Forward." *Children and Society* 18(2): 77–96.

Hodapp, Robert M, Laraine Masters Glidden, and Ann P Kaiser. 2005. "Siblings of Persons with Disabilities: Toward a Research Agenda." *Mental Retardation* 43 (November 2005): 241–254.

Hodgson, D. 1983. "Demography as Social Science and Policy Science." *Population & Development Review* 9(1): 1–34.

Holt, Louise. 2006. "Exploring 'other' Childhoods through Quantitative Secondary Analyses of Large Scale Surveys: Opportunities and Challenges for Children's Geographers." *Children's Geographies* 4(2): 143–55.

Hosmer, David W., Stanley Lemeshow, and Rodney X. Sturdivant. 2013. "Interpretation of the Fitted Logistic Regression Model." In *Applied Logistic Regression*, 47–90. John Wiley & Sons, Incorporated. https://doi.org/10.1002/0471722146.ch3.

Hubbard, Frost, Yu-chieh (Jay) Lin, Dan Zahs, and Mangyao Hu. 2016. "Sample Design." In *Cross- Cultural Survey Guidelines*, 99–149. http://unstats.un.org/unsd/hhsurveys/pdf/Household_surveys.pdf.

Hughes, Sarah M, Samuel Haddaway, and Hanzhi Zhou. 2016. "Comparing Smartphones to Tablets for Face-to-Face Interviewing in Kenya," no. 2010: 1–12.

Hunleth, Jean. 2017. Children as Caregivers: The Global Fight against Tuberculosis and HIV in Zambia. Rutgers Series in Childhood Studies.

Iliyasu, Garba, Lamaran M. Dattijo, and Abdulrazaq G. Habib. 2017. "Ebola Virus Disease and Pregnancy Outcomes: A Review of the Literature." *Tropical Journal of Obstetrics and Gynecology* 34: 6–10.

Insel, Thomas. 2013. "The Four Kingdoms of Autism." *National Institute of Mental Health*. 2013. https://www.nimh.nih.gov/about/directors/thomas-insel/blog/2013/the-four-kingdoms-of-autism.shtml.

Jalloh, Mohamed F., Paul Sengeh, Rebecca E. Bunnell, Mohammad B. Jalloh, Roeland Monasch, Wenshu Li, Jonathan Mermin, et al. 2020. "Evidence of Behaviour Change during an Ebola Virus Disease Outbreak, Sierra Leone." *Bulletin of the World Health Organization* 98 (5): 330-340B.

Journal of the American Medical Association (JAMA). "JAMA Instructions to Authors." Accessed November 28, 2019. https://jamanetwork.com/journals/jama/pages/instructions-for-authors.

James, Allison. 2007. "Giving Voice to Children's Voices: Practices and Problems, Pitfalls and Potentials." *American Anthropologist* 109(2): 261–72.

Jamieson, Denise J., Jane E. Ellis, Daniel B. Jernigan, and Tracee A. Treadwell. 2006. "Emerging Infectious Disease Outbreaks: Old Lessons and New Challenges for Obstetrician-Gynecologists." *American Journal of Obstetrics and Gynecology* 194 (6): 1546–55.

Jamieson, Denise J., Timothy M. Uyeki, William M. Callaghan, Dana Meaney-Delman, and Sonja A. Rasmussen. 2014. "What Obstetrician–Gynecologists Should Know About Ebola." *Obstetrics & Gynecology* 124 (5): 1005–10.

Jerven, Morten. 2013. Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It. Ithaca: Cornell University Press.

Johnson, Timothy P., Diane O'Rourke, Jane Burris, and Linda Owen. 2002. "Culture and Survey Nonresponse." In *Survey Nonresponse*, edited by Robert M. Groves, 55–69. Wiley.

Johnson, Timothy P, and Fons J R van de Vijver. 2003. "Social Desirability in Cross-Cultural Research." *Cross-Cultural Survey Methods*, 195–204.

Johnson, Timothy P., Beth-Ellen Pennell, Ineke A.L. Stoop, and Brita Dorer, eds. 2019. Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Context (3MC). John Wiley & Sons, Inc.

Johnson, Timothy, and Linda Owens. 2002. "Survey Response Rate Reporting in the Professional Literature." *American Association for Public Opinion Research* - Section on Survey Research Methods, 127–33.

Jones, Susan A., Somasundari Gopalakrishnan, Charles A. Ameh, Sarah White, and Nynke R. Van Den Broek. 2016. "'Women and Babies Are Dying but Not of Ebola': The Effect of the Ebola Virus Epidemic on the Availability, Uptake and Outcomes of Maternal and Newborn Health Services in Sierra Leone." *BMJ Global Health* 1 (3). https://doi.org/10.1136/bmjgh-2016-000065.

Kalton, Graham. 1983. Introduction to Survey Sampling. Newbury Park, CA: Sage Publications.

Kalton, Graham, and Ismael Flores-Cervantes. 2003. "Weighting Methods." Journal of Official Statistics 19 (2): 81–97.

Kelley, Nancy. 2006. "Children's Involvement in Policy Formation." Children's Geographies 4(1): 37–44.

Kisanji, Joseph. 1995. "Interface between Culture and Disability in the Tanzanian Context: Part I." *International Journal of Disability, Development and Education* 42 (2): 93–108.

Kish, L. 1965. Survey Sampling. New York: John Wiley & Sons, Inc.

Kohler, Hans-Peter, and Illiana Kohler. 2002. "Fertility Decline in Russia in the Early and Mid-1990s: The Role of Economic Uncertainty and Labour Market Crises." *European Journal of Population* 18 (3): 233–62.

Kondo, Michelle C, Kent DW Bream, Frances K Barg, and Charles C Branas. 2014. "A Random Spatial Sampling Method in a Rural Developing Nation" *BMC Public Health* 14 (1): 1–8.

Kovshoff, Hanna, Katie Cebula, Hsiao Wei Joy Tsai, and Richard P. Hastings. 2017. "Siblings of Children with Autism: The Siblings Embedded Systems Framework." *Current Developmental Disorders Reports* 4 (2): 37–45.

Kramer, Laurie, and Katherine J. Conger. 2009. "What We Learn from Our Sisters and Brothers: For Better or for Worse." In New Directions for Child and Adolescent Development, edited by L. Kramer and K.J. Conger, 126:1–12. San Francisco: Jossey-Bass.

Kucharski, Adam, and W. John Edmunds. 2014. "Case Fatality Rate for Ebola Virus Disease in West Africa." *The Lancet* 384 (9950): 1260–61.

Kuépié, Mathias, David Shapiro, and Michel Tenikue. 2015. "Access to Schooling and Staying in School in Selected Sub-Saharan African Countries." *African Development Review* 27 (4): 403–14.

Kumar, Naresh. 2007. "Spatial Sampling Design for a Demographic and Health Survey." *Population Research and Policy Review* 26 (5–6): 581–99.

Kviz, Frederick J. 1977. "Toward a Standard Definition of Response Rate." *The Public Opinion Quarterly* 41 (2): 265–67.

Lalwani, Ashok K., Sharon Shavitt, and Timothy Johnson. 2006. "What Is the Relation between Cultural Orientation and Socially Desirable Responding?" Journal of Personality and Social Psychology 90 (1): 165–78.

Lee Rodgers, Joseph, Craig A. St John, and Ronnie Coleman. 2005. "City Bombing? An Analysis of Births in Metropolitan Counties in Oklahoma, 1990-1999." *Demography* 42 (4): 675–92.

Leeper, Thomas J. 2019. "Where Have the Respondents Gone? Perhaps We Ate Them All." *Public Opinion Quarterly* 83 (Special Issue): 280–88. https://doi.org/10.1093/poq/nfz010.

Levison, Deborah, and Anna Bolgrien. 2020. "Using Cartoon Videos to Survey Children and Adolescents in the Global South: A Tanzanian Example." *Statistical Journal of the IAOS* 36 (S1): S147–59.

Levison, Deborah, Deborah S. DeGraff, and Esther W. Dungumaro. 2017. "Implications of Environmental Chores for Schooling: Children's Time Fetching Water and Firewood in Tanzania." *European Journal of Development Research*, 1–18.

Lévy, Yves, Clifford Lane, Peter Piot, Abdul Habib Beavogui, Mark Kieh, Bailah Leigh, Seydou Doumbia, et al. 2018. "Prevention of Ebola Virus Disease through Vaccination: Where We Are in 2018." *The Lancet*, 787–90.

Lindstrom, D P, and B Berhanu. 1999. "The Impact of War, Famine, and Economic Decline on Marital Fertility in Ethiopia." *Demography* 36 (2): 247–61.

Lipton, Jonah. 2019. "Taking Life 'Off Hold': Pregnancy and Family Formation during the Ebola Crisis in Freetown, Sierra Leone." In *Pregnant in the Time of Ebola: Women and Their Children in the 2013-2015 West African Epidemic*, edited by David Schwartz, Julienne Ngoungoung Anoko, and Sharon A. Abramowitz, 329–37. Springer.

Loeb, Mitchell E., Arne H. Eide, and Daniel Mont. 2008. "Approaching the Measurement of Disability Prevalence: The Case of Zambia." *Alter* 2 (1): 32–43.

Lovaton, Rodrigo, Aine Seitz Mccarthy, Dorothy Gondwe, Phatta Kirdruang, and Uttam Sharma. 2014. “Water, Walls and Bicycles: Wealth Index Composition Using Census Microdata.” Minnesota Population Center Working Paper Series.

Lynn, Peter, and Gerry Nicolaas. 2010. “Making Good Use of Survey Paradata.” *Survey Practice* 3 (2): 1–5.

Magnani, Robert, Keith Sabin, Tobi Saidel, and Douglas Heckathorn. 2005. “Review of Sampling Hard-to-Reach and Hidden Populations for HIV Surveillance.” *AIDS*, Supplement 19(2): 67–72.

McKay G, Black B, Mbambu Kahamba S, Wheeler E, Mearns S, Janvrin A. Not All That Bleeds Is Ebola: How has the DRC Ebola outbreak impacted Sexual and Reproductive Health in North-Kivu? New York, USA: The International Rescue Committee 2019.

Makela, Susanna, Yajuan Si, and Andrew Gelman. 2014. “Statistical Graphics for Survey Weights.” *Revista Colombiana de Estadística* 37 (2Spe): 285–95.

Mapunda, Phidea H., Abich D. Omollo, and Theodora A.L. Bali. 2017. “Challenges in Identifying and Serving Students with Special Needs in Dodoma, Tanzania.” *International Journal of Child Care and Education Policy* 11 (1).

Marlowe, Frank. 1999. “Showoffs or Providers? The Parenting Effort of Hadza Men.” *Evolution and Human Behavior* 20 (6): 391–404.

Marpsat, Maryse, and Nicolas Razafindratsima. 2010. “Survey Methods for Hard-to-Reach Populations: Introduction to the Special Issue.” *Methodological Innovations Online* 5(2): 3.1-16.

Marteleto, Letícia J., Abigail Weitzman, Raquel Zanatta Coutinho, and Sandra Valongueiro Alves. 2017. “Women’s Reproductive Intentions and Behaviors during the Zika Epidemic in Brazil.” *Population and Development Review* 43 (2): 199–227.

Marton, Krisztina, and Lowndes F Stephens. 2001. “The New York Times’ Conformity to AAPOR Standards of Disclosure for the Reporting of Public Opinion Polls.” *Journalism & Mass Communication Quarterly* 78 (3): 484–502.

Mason, Karen Oppenheim. 1997. “Explaining Fertility Transitions.” *Demography* 34 (4): 443. https://doi.org/10.2307/3038299.

Mate, Suzanne E., Jeffrey R. Kugelman, Tolbert G. Nyenswah, Jason T. Ladner, Michael R. Wiley, Thierry Cordier-Lassalle, Athalia Christie, et al. 2015. “Molecular Evidence of Sexual Transmission of Ebola Virus.” *New England Journal of Medicine* 373 (25): 2448–54.

Maulik, Pallab K., and Gary L. Darmstadt. 2007. “Childhood Disability in Low- and Middle-Income Countries: Overview of Screening, Prevention, Services, Legislation, and Epidemiology.” *Pediatrics* 120 (Supplement): S1–55. https://doi.org/10.1542/peds.2007-0043B.

Mbwilo, G.S.K., B. Smide, and C. Aarts. 2010. “Family Perceptions in Caring for Children and Adolescents with Mental Disabilities: A Qualitative Study from Tanzania.” *Tanzania Journal of Health Research* 12 (2).

McGinn, Therese, Judy Austin, Katherine Anfinson, Ribka Amsalu, Sara E. Casey, Shihab Fadulalmula, Anne Langston, et al. 2011. "Family Planning in Conflict: Results of Cross-Sectional Baseline Surveys in Three African Countries." *Conflict and Health* 5 (1): 1–8. https://doi.org/10.1186/1752-1505-5-11.

McGuire, Shirley, and Lilly Shanahan. 2006. "Sibling Experiences in Diverse Family Contexts." *Child Development Perspectives* 4 (2): 72–79.

McKendrick, John H. 2001. "Coming of Age: Rethinking the Role of Children in Population Studies." *International Journal of Population Geography* 7(6): 461–72.

McNairy Molly L, Olga Tymejczyk , Vanessa Rivera , Grace Seo , Audrey Dorélien , Mireille Peck, Jacky Petion, Kathleen Walsh, Anna Bolgrien, Denis Nash, Jean Pape, and Daniel W. Fitzgerald. 2019. "High Burden of Non-communicable Diseases among a Young Slum Population in Haiti." *J Urban Health*. Dec; 96(6):797-812.

Menard, Scott. 2002. "An Introduction to Logistic Regression Diagnostics." In *Applied Logistic Regression Analysis*, 68–91. Thousand Oaks: Sage Publications.

Menéndez, Clara, Anna Lucas, Khátia Munguambe, and Ana Langer. 2015. "Ebola Crisis: The Unequal Impact on Women and Children's Health." *The Lancet Global Health* 3 (3): e130.

Milligan, Paul, Alpha Njie, and Steve Bennett. 2004. "Comparison of Two Cluster Sampling Methods for Health Surveys in Developing Countries." *International Journal of Epidemiology* 33 (3): 469–76.

Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.2 [dataset]. Minneapolis, MN: IPUMS, 2019. https://doi.org/10.18128/D020.V7.2

Mizunoya, Suguru, Sophie Mitra, and Izumi Yamasaki. 2018. "Disability and School Attendance in 15 Low- and Middle-Income Countries." *World Development* 104: 388–403.

Montana, Livia, Peter M. Lance, Chris Mankoff, Ilene S. Speizer, and David Guilkey. 2016. "Using Satellite Data to Delineate Slum and Non-Slum Sample Domains for an Urban Population Survey in Uttar Pradesh, India." *Spatial Demography* 4 (1): 1–16.

Morgan, S. Philip, and Scott M. Lynch. 2001. "Success and Future of Demography: The Role of Data and Methods." *Annals of the New York Academy of Sciences* 954: 35–51.

Morin, R. 2002. "A 9/11 Baby Boom?" The Washington Post, 2002.

National Bureau of Statistics (NBS) Tanzania and ICF Macro. 2011. "Tanzania Demographic and Health Survey 2010." NBS and ICF Macro. Dar es Salaam, Tanzania.

National Policy on Disability, 2004. 2004. United Republic of Tanzania.

National Strategy for Inclusive Education 2018-2021. 2017.

Nobles, Jenna, Elizabeth Frankenberg, and Duncan Thomas. 2015. "The Effects of Mortality on Fertility: Population Dynamics After a Natural Disaster." *Demography* 52 (1): 15–38.

Nyangweso, Mary. 2018. "Disability in Africa: A Cultural / Religious Perspective." Peace, Human Rights Governance (PHRG), no. June. https://www.researchgate.net/publication/325642373_Disability_in_Africa_A_CulturalReligious_Perspective.

Ogechi, Nathan Oyori, and Sara Jerop Ruto. 2002. “Portrayal of Disability Through Personal Names and Proverbs in Kenya; Evidence from Ekegusii and Nandi.” *Vienna Journal of African Studies* 2 (3): 63–82.

Onwuegbuzie, Anthony J. 2007. “A Typology of Mixed Methods Sampling Designs in Social Science Research A Typology of Mixed Methods Sampling Designs in Social Science” *The Qualitative Report* 12 (2): 281–316.

OpenStreetMap Contributors. 2019. “OpenStreetMap.” https://planet.openstreetmap.org.

Opini, Bathseba, and Hezron Onditi. 2016. “Education for All and Students with Disabilities in Tanzanian Primary Schools: Challenges and Successes.” *International Journal of Educational Studies* 03 (02): 65–76.

Opperman, Sannette, and Erna Alant. 2003. “The Coping Responses of the Adolescent Siblings of Children with Severe Disabilities.” *Disability and Rehabilitation* 25 (9): 441–54.

Oran, Nicole. 2018. “Is Ebola an Outbreak or an Epidemic? What’s the Difference?” *MedCity News*. 2018.

Orsmond, Gael I., Hsin-Yu Kuo, and Marsha Mailick Seltzer. 2009. “Siblings of Individuals with an Autism Spectrum Disorder.” *Autism* 13 (1): 59–80. https://doi.org/10.1177/1362361308097119.

Peak, Corey M., Amy Wesolowski, Elisabeth zu Erbach-Schoenberg, Andrew J. Tatem, Erik Wetter, Xin Lu, Daniel Power, et al. 2018. “Population Mobility Reductions Associated with Travel Restrictions during the Ebola Epidemic in Sierra Leone: Use of Mobile Phone Data.” *International Journal of Epidemiology* 47 (5): 1562–70.

Persons with Disabilities Act, 2010. 2010. United Republic of Tanzania.

Pilowsky, Tammy, Nurit Yirmiya, Osnat Doppelt,Varda Gross-Tsur and Ruth S. Shalev. 2004. “Social and Emotional Adjustment of Siblings of Children with Autism.” *Journal of Child Psychology and Psychiatry* 45 (4): SP-855-65.

Possi, Mwajabu K., and Joseph R. Milinga. 2017. “Special and Inclusive Education in Tanzania: Reminiscing the Past, Building the Future.” *Educational Process: International Journal* 6 (4): 55–73.

Priebe, Jan. 2018. “Disability and Its Correlates in a Developing Country Context: Evidence from Multiple Datasets and Measures.” *Journal of Development Studies* 54 (4): 657–81.

Rangel, Marcos A., Jenna Nobles, and Amar Hamoudi. 2020. “Brazil’s Missing Infants: Zika Risk Changes Reproductive Behavior.” *Demography* 57 (5): 1647–1680.

Ravindran, Neeraja, and Barbara J. Myers. 2012. “Cultural Influences on Perceptions of Health, Illness, and Disability: A Review and Focus on Autism.” *Journal of Child and Family Studies* 21 (2): 311–19. https://doi.org/10.1007/s10826-011-9477-9.

Reierson Draugalis, Jolaine, Stephen Joel Coons, and Cecilia M Plaza. 2008. “Best Practices for Survey Research Reports: A Synopsis for Authors and Reviewers.” *American Journal of Pharmaceutical Education* 72 (1): 1–8.

Richards, Paul, Joseph Amara, Mariane C. Ferme, Prince Kamara, Esther Mokuwa, Amara Idara Sheriff, Roland Suluku, and Maarten Voors. 2015. “Social Pathways for Ebola Virus

Disease in Rural Sierra Leone, and Some Implications for Containment." *PLOS: Neglected Tropical Diseases* 9 (4).

Ritter, Joseph, Monique Borgerhoff Mulder, Kari Hartwig, Susan James, Deborah Levison, Ester Ngadaya, and Craig Packer. 2010. "The Whole Village Project: A Platform for Evaluating Rural Development Projects". Minnesota Population Center Working Paper Series. Minneapolis.

Rohwerder, Brigitte. 2014. "Impact and Implications of the Ebola Crisis." www.gsdrc.org.

Roos, Robert. 2014. "Migrations in West Africa Seen as Challenge to Stopping Ebola." November 14, 2014. cidrap.umn.edu/news-perspective/2014/11/migrations-west-africa-seen-challenge-stopping-ebola.

Roper, Susanne Olsen, Diane W Allred, Barbara Mandleco, Donna Freeborn, and Tina Dyches. 2014. "Caregiver Burden and Sibling Relationships in Families Raising Children with Disabilities and Typically Developing Children" *Families, Systems, and Health* 32 (2): 241–46.

Rosa, Edinete Maria, and Jonathan Tudge. 2013. "Urie Bronfenbrenner's Theory of Human Development: Its Evolution from Ecology to Bioecology." *Journal of Family Theory & Review* 5 (4): 243–58.

Ross, Penelope, and Monica Cuskelly. 2006. "Adjustment, Sibling Problems and Coping Strategies of Brothers and Sisters of Children with Autistic Spectrum Disorder." *Journal of Intellectual and Developmental Disability* 31 (2): 77–86.

Rossetti, Zach, and Sarah Hall. 2015. "Adult Sibling Relationships with Brothers and Sisters with Severe Disabilities." *Research and Practice for Persons with Severe Disabilities* 40 (2): 120–37.

Rossiter, Lucille, and Donald Sharpe. 2001. "The Siblings of Individuals with Mental Retardation: A Quantitative Integration of the Literature." *Journal of Child and Family Studies* 10 (1): 65–84.

Sadler, Georgia Robins, Hau Chen Lee, Rod Seung Hwan Lim, and Judith Fullerton. 2010. "Recruitment of Hard-to-Reach Population Subgroups via Adaptations of the Snowball Sampling Strategy." *Nursing and Health Sciences* 12 (3): 369–74.

Salganik, Matthew J. 2006. "Variance Estimation, Design Effects, and Sample Size Calculations for Respondent-Driven Sampling." *Journal of Urban Health* 83 (7 SUPPL.): 98–112.

Saltelli, Andrea. 2020. "Ethics of Quantification or Quantification of Ethics?" *Futures* 116 (October 2019): 102509.

Sanchez, Morgan. 2016. "The Sibling Disability Experience: An Analysis of Studies Concerning Non- Impaired Siblings of Individuals with Disabilities from 1960 to 1990." In *Sociological Looking at Disability: What Did We Know and When Did We Know It?* edited by Sharon N. Barnartt Sara E. Green, (Research, 241–59. Emerald Group Publishing Limited.

Savel, Craig, Stan Mierzwa, Pamina Gorbach, Michelle Lally, Gregory Zimet, Kristin Meyer, and Sameer Souidi. 2014. "Web-Based, Mobile-Device Friendly, Self-Report Survey System Incorporating Avatars and Gaming Console Techniques." *Online Journal of Public Health Informatics* 6 (2): 1–11.

Scelfo, J. 2002. “A 9-11 Baby Boomlet.” *Newsweek*, 2002.

Schooler, Jonathan W. 2014. “Metascience Could Rescue the ‘Replication Crisis.’” *Nature* 515 (7525): 9.

Schuntermann, Peter. 2007. “The Sibling Experience: Growing up with a Child Who Has Pervasive Developmental Disorder or Mental Retardation.” *Harvard Review of Psychiatry* 15 (3): 93–108.

Schwartz, David, Sharon A. Abramowitz, and Julienne Ngoungoung Anoko. 2019. Pregnant in the Time of Ebola: Women and Their Children in the 2013-2015 West African Epidemic. Global Maternal and Child Health: Medical, Anthropological, and Public Health Perspectives. Cham, Switzerland: Springer Nature Switzerland AG.

Seisay, Alhssan Lans, and Mohamed Koblo Kamara. 2017. “Sierra Leone 2015. Population and Housing Census. Thematic Report on Mortality.” Report for Statistics Sierra Leone.

Shandra, Carrie L., Roger C. Avery, Dennis P. Hogan, and Michael E. Msall. 2012. “Child and Adult Disability in the 2000 Census: Disability Is a Household Affair.” *Disability Health Journal* 5 (4): 1–16.

Shannon, Harry S., Royce Hutson, Athena Kolbe, Bernadette Stringer, and Ted Haines. 2012. “Choosing a Survey Sample When Data on the Population Are Limited: A Method Using Global Positioning Systems and Aerial and Satellite Photographs.” *Emerging Themes in Epidemiology* 9: 1–7.

Shultz, James M., Janice L. Cooper, Florence Baingana, Maria A. Oquendo, Zelde Espinel, Benjamin M. Althouse, Louis Herns Marcelin, et al. 2016. “The Role of Fear-Related Behaviors in the 2013–2016 West Africa Ebola Virus Disease Outbreak.” *Current Psychiatry Reports* 18 (11).

Sida. 2014. “Disability Rights in Tanzania.” Sweden. www.sida.se.

Skotko, Brian G., and Susan P. Levine. 2006. “What the Other Children Are Thinking: Brother and Sisters of Persons with Down Syndrome.” *American Journal of Medical Genetics* 142 (C): 180–86. https://doi.org/10.1002/ajmg.c.

Smith-Greenaway, Emily. 2015. “Educational Attainment and Adult Literacy: A Descriptive Account of 31 Sub-Saharan Africa Countries.” *Demographic Research* 33 (1): 1015–34.

Smyth, Jolene D., Kristen Olson, and Mathew Stange. 2019. “Within-Household Selection Methods: A Critical Review and Experimental Examination.” *Experimental Methods in Survey Research*, 23–45.

Sobotka, Tomáš, Vegard Skirbekk, and Dimiter Philipov. 2011. “Economic Recession and Fertility in the Developed World.” *Population and Development Review* 37 (2): 267–306.

Söder, Mårten. 1989. “Disability as a Social Construct: The Labelling Approach Revisited.” European *Journal of Special Needs Education* 4 (2): 117–29. https://doi.org/10.1080/0885625890040204.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. “What Are We Weighting For?” *Journal of Human Resources* 50 (2): 301–16. https://doi.org/10.3368/jhr.50.2.301.

Sonik, Rajan A., Susan L. Parish, and Eliana S. Rosenthal. 2016. “Sibling Caregivers of People with Intellectual and Developmental Disabilities: Sociodemographic Characteristics and

Material Hardship Prevalence." *Intellectual and Developmental Disabilities* 54 (5): 332–41. https://doi.org/10.1352/1934-9556-54.5.332.

Staff, Reuters. 2014. "Sierra Leone Shuts Borders, Closes Schools to Fight Ebola." *Reuters*, June 11, 2014. http://uk.reuters.com/article/2014/06/11/us-health-ebola-leone-idUKKBN0EM2CG20140611.

Stedman, Richard C, Nancy A Connelly, Thomas A Heberlein, Danial J. Decker, and Shorna B Allred. 2019. "The End of the (Research) World as We Know It? Understanding and Coping with Declining Response Rates to Mail Surveys." *Society & Natural Resources* 32 (10): 1139–54. https://doi.org/10.1080/08941920.2019.1587127.

Stone-MacDonald, Angi, and Gretchen Butera. 2012. "Cultural Beliefs and Attitudes about Disability in East Africa." Review of Disability Studies: An International Journal 8 (1).

Stone-Macdonald, Angi. 2012. "Cultural Beliefs about Disability in Practice: Experiences at a Special School in Tanzania." *International Journal of Disability, Development, and Education* 59 (4): 393–407. https://doi.org/10.1080/1034912X.2012.723947.

Stoneman, Zolinda, and Phyllis Waldman Berman. 1993. The Effects of Mental Retardation, Disability, and Illness on Sibling Relationships: Research Issues and Challenges. Baltimore: Paul H Brooks Publishing Co.

Stoneman, Zolinda. 2001. "Supporting Positive Sibling Relationships During Childhood." *Mental Retardation and Developmental Disabilities Research Revie*ws 142: 134–42.

Stoneman, Zolinda. 2005. "Siblings of Children with Disabilities: Research Themes." *Mental Retardation* 43 (5): 339–50.

Survey Research Center. 2016. "Guidelines for Best Practice in Cross-Cultural Surveys." Ann Arbor, MI: Retrieved June 19, 2020.

Susman, Joan. 1994. "Disability and Stigma." *Social Science and Medicine* 38 (1): 15–22. https://doi.org/10.3917/cm.094.0125.

Taft, Jessica K. 2019. The Kids Are in Charge: Activism and Power in Peru's Movement of Working Children. NYU Press.

Teitler, Julien O., Nancy E. Reichman, and Susan Sprachman. 2003. "Costs and Benefits of Improving Response Rates for a Hard-to-Reach Population." *Public Opinion Quarterly* 67(1): 126–38.

Thiede, Brian C., Matthew Hancock, Ahmed Kodouda, and James Piazza. 2020. "Exposure to Armed Conflict and Fertility in Sub-Saharan Africa." *Demography* 57 (6): 2113–41. https://doi.org/10.1007/s13524-020-00923-2.

Thompson, Stephen. 2017. "Disability Prevalence and Trends." *K4D Help Desk Report*.

Tisdall, E. Kay M., and Samantha Punch. 2012. "Not so 'New'? Looking Critically at Childhood Studies." *Children's Geographies* 10(3): 249–64.

Training, Ministry of Education and Vocational. 2009. National Strategy on Inclusive Education 2009-2017.

Trinitapoli, Jenny, Sara Yeatman, and Jasmine Fledderjohann. 2014. "Sibling Support and the Educational Prospects of Young Adults in Malawi." *Demographic Research* 30 (1): 547–78. https://doi.org/10.4054/DemRes.2014.30.19.

Turner, Anthony G., Robert J. Magnani, and Muhammad Shuaib. 1996. "A Not Quite as Quick but Much Cleaner Alternative to the Expanded Programme on Immunization (EPI) Cluster Survey Design." *International Journal of Epidemiology* 25 (1): 198–203. https://doi.org/10.1093/ije/25.1.198.

Underwood, J H. 1984. "Effects of the 1918 Influenza Pandemic Mortality Experience on Subsequent Fertility of the Native Population of Guam." *Micronesia* 19 (1–2): 1–10.

UNICEF. 2017. "Education Fact Sheet." Tanzania National Bureau of Statistics.

UNICEF. 2020. "Choosing the Sample." In Multiple-Indicator Survey Handbook. New York. Accessed March 11, 2021. http://mics.unicef.org/files?job=W1siZiIsIjIwMTUvMDQvMDMvMDYvNDIvNDgvMjg2L2NoYXAwNC5wZGYiXV0&sha=d31cdb905d60500d.

United Nations. 1990. Convention on the Rights of the Child.

United Nations. 2006. "Convention on the Rights of Persons with Disabilities."

United Republic of Tanzania President's Office. 2016. "Pre-Primary, Primary, and Secondary Education Statistics in Brief." Dodoma.

Van der Merwe, Christine, Juan Bornman, Dana Donohue, and Michal Harty. 2017. "The Attitudes of Typically Developing Adolescents towards Their Sibling with Autism Spectrum Disorder." *South African Journal of Communication Disorders* 64 (1): 1–7.

Verardi, Sabrina, Donatien Dahourou, Jennifer Ah-Kion, Uma Bhowon, Caroline Ng Tseung, Denis Amoussou-Yeye, Marcel Adjahouisso, et al. 2010. "Psychometric Properties of the Marlowe-Crowne Social Desirability Scale in Eight African Countries and Switzerland." *Journal of Cross-Cultural Psychology* 41 (1): 19–34.

Vermaes, Ignace P R, Anna M J Van Susante, C Ms, and Hedwig J A Van Bakel. 2012. "Psychological Functioning of Siblings in Families of Children with Chronic Health Conditions: A Meta-Analysis." *Journal of Pediatric Psychology* 37 (2): 166–84.

Vu, Alexander, Nhan Tran, Kiemanh Pham, and Saifuddin Ahmed. 2011. "Reliability of the Marlowe-Crowne Social Desirability Scale in Ethiopia, Kenya, Mozambique, and Uganda*." BMC Medical Research Methodology* 11 (1): 162. https://doi.org/10.1186/1471-2288-11-162.

Vyas, Seema, and Lilani Kumaranayake. 2006. "Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis." *Health Policy and Planning* 21 (6): 459–68.

Vygen, Sabine, Amanda Tiffany, Monica Rull, Alexandre Ventura, Anja Wolz, Amara Jambai, and Klaudia Porten. 2016. "Changes in Health-Seeking Behavior Did Not Result in Increased All-Cause Mortality during the Ebola Outbreak in Western Area, Sierra Leone." *American Journal of Tropical Medicine and Hygiene* 95 (4): 897–901.

Wadsworth, Anne. 2015. "The State of Education in Tanzania." *Girls Education Collaborative*. 2015. https://gec.ngo/the-state-of-education-in-tanzania/.

Wampler, Peter J., Richard R. Rediske, and Azizur R. Molla. 2013. “Using ArcMap, Google Earth, And Global Positioning Systems to Select and Locate Random Households in Rural Haiti.” *International Journal of Health Geographics* 12. https://doi.org/10.1186/1476-072X-12-3.

Washington Group on Disability Statistics. 2020. “The Washington Group Short Set Enhanced (WG-SS Enhanced).”

Washington Group. 2018. “Regional Guidebook to Improve Disability Data Collection and Analysis in the Arab Countries Implementing the Washington Group Questions on Functioning.” Beirut.

Watkins, Susan Cotts. 1993. “If All We Knew About Women Was What We Read in Demography, What Would We Know? Published by: *Population Association of America*

Watters, John K, and Patrick Biernacki. 1989. “Targeted Sampling: Options for the Study of Hidden Populations.” *Social Problems* 36 (4): 416–30.

Weisner Thomas S. 1989. “Comparing sibling relationships across cultures.” In *Sibling Interaction across Cultures: Theoretical and Methodological Issues*, edited by Patricia Goldring Zukow. New York: Springer-Verlag.pp. 11–22.

Weisner, Thomas S. 1993. “Ethnographic and Ecocultural Perspectives on Sibling Relationships.” In *The Effects of Mental Retardation, Disability, and Illness on Sibling Relationships: Research Issues and Challenges*, edited by Zolinda Stoneman and Phyllis Waldman Berman, 51–83. Paul H Brooks Publishing Co.

Whyte, Susan Reynolds. 1991. “Attitudes towards Mental Health Problems in Tanzania” *Acta Psychiatrica Scandinavica*., Suppl. 364. Vol. 83.

Winters, Maike, Helena Nordenstedt, and Helle Mölsted Alvesson. 2020. “Reporting in a Health Emergency: The Roles of Sierra Leonean Journalists during the 2014-2015 Ebola Outbreak.” *PLoS Neglected Tropical Diseases* 14 (5): 1–14. https://doi.org/10.1371/journal.pntd.0008256.

World Bank Group. 2015. “The Socio-Economic Impacts of Ebola in Sierra Leone: Results from High Frequency Cell Phone Survey.”

World Health Organization. 2014. “Emergencies Preparedness, Response Sierra Leone: A Traditional Healer and a Funeral,” 11–13.

World Health Organization. 2016. “Ebola Virus Disease.” *Situation Report*, no. JUNE: 1–2. http://apps.who.int/iris/bitstream/10665/208883/1/ebolasitrep_10Jun2016_eng.pdf?ua=1.

World Health Organization. 2017. “Ebola Virus Disease: Fact Sheet.” http://www.who.int/mediacentre/factsheets/fs103/en/.

Xie, Yu. 2000. “Demography: Past, Present, and Future.” *Journal of the American Statistical Association* 95(450): 670–73.

Yansaneh, Ibrahim S. 2003. “Construction and Use of Sample Weights.” *Designing of Household Sample Surveys*, no. November: 1–12.

Zimmerman, Linnea. 2017. “PMA2020 General Sampling Strategy.” https://www.pma2020.org/reports/pma2020-sdp-sampling-memo-en.

Zola, Irving K. 1993. “Disability Statistics, What We Count and What It Tells Us.” *Journal of Disability Policy Studies* 4 (2): 9–39. https://doi.org/10.1177/104420739300400202.

Zukow, Patricia Goldring. 1989. “Communicating Across Disciplines: On Integrating Psychological and Ethnographic Approaches to Sibling Research.” In *Sibling Interaction across Cultures: Theoretical and Methodological Issues*, edited by Patricia Goldring Zukow. New York: Springer-Verlag.

Appendix

School enrollment for children with disabilities (Footnote 58)

It is well established in the literature that children with disabilities in Tanzania are less likely than their peers without disabilities to be enrolled in school (Opini and Onditi 2016). Much of the information calculating the enrollment of students with disabilities is constructed from census data and national reports. In this Appendix I, I estimate the following logistic regression model for children 7 to 17 who are living at home and are the child of the head of the household.

$$\ln\left[\frac{p}{1-p}\right] = \alpha + \beta_1 * X_1 + \beta_2 * X_2 + \cdots + \beta_k * X_k$$

where *p* is the probability that a child is currently enrolled in school, the main independent variables (X_1) is disability status. Control variables (X) include the same demographic and household characteristics found in the main results: age, sex, number of siblings, wealth index, and urban status. The intercept is denoted (α). In this model, I allow "only children" to be included as this analysis is not concerned with sibling status.

Table A1 shows the odds ratios for school enrollment comparing children age 7 to 17 who have disabilities to children who do not have disabilities. Children with disabilities are 68% less likely (odds ratio = 0.33 with no controls; odds ratio = 0.324 with demographic and household controls) to be enrolled in school compared to children without disabilities. The coefficients for the control variables are all significant and similar to direction and magnitude to the analysis of sibling status for children without disabilities.

Table A1: Odds Ratios for School Enrollment for Children 7-17 with Disabilities

	(1)	(2)	(3)
Child has disability			
(ref. no disability)	0.3321***	0.3256***	0.3244***
	(0.01)	(0.01)	(0.01)
Age		0.9578***	0.9408***
		(0.00)	(0.00)
Sex		1.1575***	1.1450***
		(0.01)	(0.01)
N Children in HH		0.8748***	0.9262***
		(0.00)	(0.00)
Quantiles			
(Ref. = Lowest)			
2			1.5448***
			(0.01)
3			2.5620***
			(0.03)
4			3.7261***
			(0.04)
5 (Highest)			8.1803***
			(0.14)
Urban/ Rural			
(ref. Urban)			1.7982***
			(0.02)
N	882299	882299	882299

Note: *p<0.10 **p< 0.05 *** p<0.010

Age sensitivity of school enrollment and sibling relationships

I also tested the sensitively to the age cut-offs I use in the data in three ways. First, I lowered the age of enrollment to include potential pre-primary students age five and six. Next, I included all children under the age of 17 without disabilities to be my unit of interest. Finally, I allowed the age of the siblings with and without disabilities to be greater than 17 years old. This

final specification allows the potential for siblings who are living at home who are older than 17 years old [See footnote 56]. In many families with persons with disabilities, it is likely the individual with a disability will remain at home, but the presence of this individual with disabilities may still require younger siblings to assist in care or other household responsibilities. However, none of the variations on age changed the odds of school enrollment based on sibling status compared to the preferred specification.