11-24-2023

Effects of Water Access on Childhood Linear Growth in Zimbabwe

Robert Ntozini
Walden University

Walden University

College of Health Sciences and Public Policy

This is to certify that the doctoral study by

Robert Ntozini

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. W. Sumner Davis, Committee Chairperson, Public Health Faculty
Dr. Peter Anderson, Committee Member, Public Health Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2023

Abstract

Effects of Water Access on Childhood Linear Growth in Zimbabwe

by

Robert Ntozini

MPH, University of South Africa, 2011

BS, University of Zimbabwe, 1992

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Public Health

Walden University

November 2023

Abstract

Worldwide an estimated 149.2 million children under 5 years were stunted in 2020, and 58% lived in low and middle-income (LMIC) countries characterized by poor nutrition and poor access to water, sanitation, and hygiene (WaSH) services. The best nutrition interventions implemented in LMIC can reduce childhood stunting by up to a third of the height deficit; however, evidence that WaSH interventions may also reduce stunting is mixed. The objective of this study was to investigate the effects of water access on childhood linear growth and stunting. Generalized estimating equations framework was used to estimate associations between water access during pregnancy and two outcomes: length-for-age z-score (LAZ) and childhood stunting at 18 months of age using secondary data from a cluster-randomized clinical trial in a rural setting. 4,036 mother/infant dyads drawn from 209 clusters were matched on water access and child linear growth. HIV and stunting prevalence were 16% and 33%, respectively; 62% of the households used water from boreholes, and ~1% had piped water in their homes. Water volume per capita per day was associated with improved child linear growth, mean ΔLAZ = 0.10 (95% CI: 0.004; 0.20), $p = 0.04$; the association was stronger among children born to HIV-infected mothers but not among HIV-unexposed children. Distance and one-way walk time to a primary water source were weakly associated with linear growth, but water quality was not. Community-level water access measures were not consistently associated with mean LAZ; volume of water per person per day predicted child linear growth, and the impact was greater among children born to HIV-infected mothers. Implications for positive social change include precision targeting of vulnerable subpopulations, which may be a more effective approach to addressing the current child linear growth challenges.

Effects of Water Access on Childhood Linear Growth in Zimbabwe

by

Robert Ntozini

MPH, University of South Africa, 2011

BS, University of Zimbabwe, 1992

Doctoral Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Public Health

Walden University

November 2023

Dedication

I dedicate this work to my creator, The Almighty God, for the wisdom, knowledge, and inspiration to undertake this research. I also dedicate this work to my family: my loving wife Rosemary Paidemoyo, and my children Siphosenkosi Kupakwashe, Sibusisiwe, Zanele and Zinhle, who all endured many lonely days and nights but still encouraged and inspired me to carry on with this work. I dedicate this work to my parents, my late father Petros Nqanyana Ntozini and my dear mother Rebecca Nqanyana Ntozini for their love and wisdom, and for sacrificing many things in their lives for my education. I dedicate this work to all children, especially those living in the two districts of Chirumanzu and Shurugwi in Zimbabwe, and those living in low and middle-income countries, who all suffer the burden of stunting. Finally, I dedicate this work to all public health workers the world over who struggle to make a difference to improving the lives of children to make them attain their full developmental potential.

Acknowledgments

My deepest appreciation for this work goes to my family; my loving wife Rosemary Paidemoyo Ntozini, and my dear children Siphosenkosi Kupakwashe Ntozini, Sibusisiwe Ntozini, Zanele Ntozini and Zinhle Ntozini. You all endured many lonely days and nights but still encouraged and inspired me to carry on with this work.

I appreciate my parents; my father the late Mr Petros Nqanyana Ntozini and my dear mother Mrs Rebecca Nqanyana, for their love and wisdom and for sacrificing many finer things in their lives for my education and wellbeing. I am grateful to you.

I appreciate my dissertation committee; the Chair Dr. W. Sumner Davis, the second member Dr. Peter Anderson, and finally the Program Director Dr. Holly Godwin. I thank you all for providing very critical feedback during the development of this capstone and helping me develop a stronger research plan.

Finally, I appreciate all my lecturers in the DrPH program at Walden University who taught me in various courses and contributed to my understanding of public health. Their inputs equipped me with the skills needed to make this study a possibility.

Table of Contents

List of Tables

List of Figures

Section 1: Foundation of the Study and Literature Review

Globally 149.2 million children under 5 years were stunted in 2020 according to the United Nations Children's Fund (UNICEF), World Health Organization (WHO) and World Bank Joint Child Malnutrition Estimates Group (WHO, 2023). The majority of stunted children live in LMIC, where a total of 135.7 million children comprising 58% of stunted children worldwide live; 57.5 million of them lived in Africa and 78.2 million in Asia (UNICEF et al., 2020). This shows that the burden of childhood stunting is disproportionately among children who live in developing countries. Over the last decade the prevalence of childhood stunting has been decreasing globally; however, the population of stunted children in sub-Saharan Africa (SSA) has been increasing with growth of the population in the region (de Onis & Branca, 2016; UNICEF et al., 2020).

There is strong evidence that childhood stunting is associated with poor access to water, sanitation, and hygiene (WaSH), which play a key role in preventing the spread of infectious diseases that can affect child growth and development (de Onis & Branca, 2016; Schmidt, 2014; Spears, 2013). WaSH remains a key determinant of child health according to a conceptual framework on childhood stunting, which was first proposed by UNICEF and WHO in 1990 and is still being used to explain the underlying causes of childhood stunting. Improvements in access to WaSH have been observed to result in reductions in child stunting, as witnessed in the case of Brazil over the last three decades (Rasella, 2013). Therefore, interventions for reducing childhood stunting should include some components of WaSH to be more effective.

One of the biggest disparities between rural and urban Zimbabwean communities is their access to water. In 2012, according to the Progress on Drinking-

Water and Sanitation Report (WHO & UNICEF, 2014), an estimated 19% of rural Zimbabweans did not have access to improved drinking-water compared to only 1% in urban areas. The Joint Monitoring Program (JMP) report by WHO and UNICEF (2014) also estimated that 34% of the Zimbabwean population in 2012 had access to piped water on premises, and the rest, 46%, 15%, and 5% had access to improved wells, unimproved wells, and surface water, respectively. Piped water into premises is only available in Zimbabwean urban centers, whereas rural water access is mainly from wells, boreholes, and even open water bodies. The median distance to water points in rural Chirumanzu and Shurugwi Districts in the Midlands Province of Zimbabwe was about 361m ranging from 146–848m, and there was a wide seasonal variation in access throughout the year (Ntozini et al., 2015). This suggests that improvements in WaSH access in rural Zimbabwe may have a significant role in the reduction of childhood stunting if water access was improved.

In this study, I explored the relationship between childhood stunting and access to water in the two rural districts of Zimbabwe in the Midlands Province. This research fills an important gap in understanding how water-access could affect childhood stunting in these rural communities. Findings from this study could have a positive social change impact on rural communities by influencing how water services may be provided to communities in resource-limited settings in order to impact child health outcomes. The impact of these findings could also influence public health policy on malnutrition and disease prevention in rural areas in LMIC.

The foundation section is divided into logical subsections: a brief literature on childhood stunting, its consequences, and the link with WaSH including water access is discussed in the introduction and background subsections. The public health

problem addressed in this study and the purpose of this research are discussed under the problem statement and purpose of the study. The research questions and hypotheses are then discussed followed by a theoretical framework that guides the design of the study and potential pathways to impact. The rest of the section discusses the selection of the study design under the nature of the study; the sources of data, and definition of terms used; assumptions and scope as well as the significance of the study.

Background

Stunting, also known as growth retardation, is a phenomenon where a child is too short for their stature at a given age. Childhood stunting is technically defined as having a length-for-age z-score (LAZ) or height-for-age z-score (HAZ) less than two standard deviations below the median length or height on the WHO child growth standard (WHO, 2016a). Children below 24 months of age are measured while lying down as length, and older children > 24 months are measured while standing up as height (WHO, 2016a). Current evidence suggests that stunting begins early in life, starting in-utero during fetal development, and continues until between 18 to 24 months infant age after that stunted children are unlikely to recover to their full potential (Victora et al., 2010).

Childhood stunting is an important indicator of the overall health status of children. Stunting is an indicator of a child's overall well-being and an accurate reflection of social inequalities (de Onis & Branca, 2016). Several researchers have shown that childhood stunting is associated with multiple public health problems that affect their lives from infancy, persist into their adulthood and even affects their offspring. According to work by various researchers, the consequences of stunting are

summarized as follows:

- Stunted children are associated with poor early childhood cognitive development outcomes,
- Stunted children are associated with attending less schooling and are likely to perform poorly in class,
- In their adult life, stunted children tend to work in less paying jobs and have reduced economic productivity,
- Stunted children are likely to become more obese in their adult life and are more susceptible to non-communicable diseases such as hypertension and heart diseases, and
- Mothers who were stunted as children are more likely to bear stunted offspring perpetuating the stunting cycle. (Adair et al., 2013; Black et al., 2008; Dewey & Begum, 2011; Stein et al., 2005; Victora et al., 2008; Vollmer et al., 2014).

This shows the significance of childhood stunting as an important public health problem and highlights the urgency of finding public health interventions that can effectively reduce the problem. Due to the complexity and far reaching effects of the problem of childhood stunting, action to reduce stunting requires a multi-pronged approach involving improvements in multiple domains including WaSH, food and nutrition security, education, better health, and reduction in poverty by raising the economic and social status of women (de Onis & Branca, 2016). Therefore, intervention approaches that target multiple pathways to child stunting may have the best chances of tackling the problem of reducing stunting.

It is estimated that in 2013 between 36–42% of children living in developing

countries in SSA and Asia were stunted (Black et al., 2013). In Zimbabwe, for example, one of the SSA countries that has high levels of childhood stunting, in 2016 there was an estimated 35% of children aged under 5 years either stunted, wasted, or underweight (WHO, 2016b). It was also reported that routine assessment of child height in addition to weight at postnatal visits was only introduced in 2012 on the "road to health card"—the postnatal child health card issued to all children at birth in Zimbabwe and used to record child growth and development ("Government Unveils New Child Health Card, Vaccine," 2012). Furthermore, only health care staff at public health facilities—where anthropometric equipment is available—can complete the card while most private health providers generally leave the new sections of the updated card blank. This means that even though childhood stunting could be monitored at child-level, this is not always happening due to other operational constraints in the health care systems in Zimbabwe. The Global Nutrition Profile for Zimbabwe (2014) reported that there are wide disparities in the absolute prevalence of childhood stunting in Zimbabwe. The report also shows that there were disparities in changes in the rates of stunting over time between the rich and the poor. For example, in 2011 while the country's average under-5 stunting prevalence was 32%, only 25% of children from the wealthiest quantile were stunted; meanwhile, 36% of children from the poorest quantile were stunted. This also shows that children from families with low socioeconomic status, who are mainly rural children, are at a higher risk of stunting.

Generally, childhood stunting has remained a hidden public health problem that is understood and discussed mostly between academics more than by public health practitioners and the affected populations. Childhood stunting often goes

unrecognized in communities where short stature is the norm, and that linear growth is normally not assessed in most primary healthcare settings (de Onis & Branca, 2016). Communities with the greatest need for interventions to reduce stunting are not aware of this as a problem and children are not getting the attention they need in these communities (Said-Mohamed et al., 2015). This suggests that countries that have the greatest need to reduce child stunting may not even take it as a serious public health problem even though it has serious detrimental consequences to their populations.

Research into childhood stunting over the last two decades focused mainly on understanding the determinants and consequences of childhood stunting. In 1990, WHO and UNICEF proposed a comprehensive conceptual framework on childhood malnutrition that depicts the basic, underlying, and immediate causes of childhood malnutrition and stunting (UNICEF & WHO, 1990). The problems and consequences of childhood stunting have since been escalated to the United Nations (UN) and WHO platforms and world leaders met in the year 2000 and set some aspirational goals to reduce childhood stunting, among other health outcomes. The Millennium Development Goals (MDGs) in 2000 as well as the recent Sustainable Development Goals (SDGs) in 2015 were targets set to improve child survival and achieve optimum growth and development (UN, n.d.-a, n.d.-b). These efforts have seen a steady global decline in stunting prevalence from 32.6% in 2000 to 22.2% by 2017; however, in SSA the population of stunted children has been increasing over the same period due to population growth (de Onis & Branca, 2016; Said-Mohamed et al., 2015; UNICEF et al., 2018). This means that even though in some regions stunting may be declining, in SSA and Asia the problem persists.

The underlying causes of childhood stunting act along different pathways and

at different levels. There are two prominent pathways to stunting in the framework proposed by UNICEF: One pathway acts through inadequate dietary intake, and the other pathway is due to disease. The two pathways also interact with disease modifying dietary intake and lake of adequate diet making children more vulnerable to disease (UNICEF & WHO, 1990). Nutrition interventions target the nutritional pathway and attempt to fill in the nutrient gap to improve child growth (Stewart et al., 2013). However, many studies have shown that good complementary feeding interventions may only have a small impact on childhood stunting for children in LMICs (Panjwani & Heidkamp, 2017). Nutrition interventions alone are inadequate for eliminating childhood stunting. A number of researchers have reported that a large part of the disease burden, including childhood stunting, is associated with having poor access to water and inadequate sanitation and hygiene (Bartram & Cairncross, 2010; Cumming & Cairncross, 2016; Prüss-Ustün et al., 2014). Countries in the world worst affected by stunting had correspondingly poor rates of water and sanitation coverage, and this association is especially significant in LMIC. Childhood stunting has been positively associated with improvements in water coverage (Esrey et al., 1991; Fry et al., 2010; Rasella, 2013). It is therefore important to examine the level of water coverage required to significantly affect child health outcomes.

Access to water can be assessed in many dimensions; proximity to water source, quantity of water used, and the quality of the water used are some dimensions used among other measures. It is important to establish which particular dimension or dimensions of access to water are the most important to prevent childhood stunting. The WHO has not provided sufficient guidance on the quantity of domestic water that is required to promote good health (Howard et al., 2020). The availability of

freshwater and distance travelled to fetch water adversely affects child health outcomes including childhood stunting (Abubakar et al., 2012; Pickering & Davis, 2012). In terms of amount of water per person per day needed to prevent morbidity and mortality in disaster settings, evidence has shown that increasing the amount of water used by each person per day reduced the risk of adverse health outcomes (De Buck et al., 2015). Water quality is another dimension of water access that may be important for childhood stunting. Improving the quality of water, by disinfection and practicing hygiene by providing soap, has shown small benefits to linear growth for children < 5 years (Dangour et al., 214). However, the quality of the evidence in the studies reviewed was poor due to weak methodologies or size of studies was too small. This highlighted a gap where more and better quality evidence is required to quantify the health benefits associated with improving water quality in resource-limited settings in order to improve child linear growth and reduce stunting.

Problem Statement

UNICEF, WHO, and the World Bank Joint Child Malnutrition Estimates Group estimated that globally 149.2 million children under 5 years were stunted in 2020 (WHO, 2023). In the same period a total of 135.7 million children comprising 58% of stunted children lived in LMICs; 57.5 million of them lived in Africa and 78.2 million in Asia (UNICEF et al., 2020). Childhood stunting is a multidimensional problem involving many interacting factors; nutrition plays a key role, but the environment in which children grow also plays a pivotal role (UNICEF & WHO, 1990). The UNICEF conceptual framework on childhood malnutrition and stunting places access to WaSH as an alternative pathway to childhood stunting mediated through diseases (UNICEF & WHO, 1990; Stewart et al., 2013). LMICs, including

Zimbabwe, have both poor WaSH coverage and high prevalence of stunting (Spears, 2013; Schmidt, 2014). According to The Global Nutrition Profile (2014) for Zimbabwe, the highest prevalence of childhood stunting is among the rural and poorest who also have the poorest access to water. Only 16.5% of rural Zimbabweans in Chirumanzu and Shurugwi Districts walked less 100m to fetch drinking water (Ntozini et al., 2015).

A number of studies quantified the health benefit of improving access to water; however, they reported mixed results, and the quality of the evidence in some of these studies was poor (Esrey et al., 1991; Fry et al., 2010; Overbo et al., 2016; Rasella, 2013). For example, Harris et al. (2017) argued that individual-level access to WaSH resources might be less important than community-level coverage. Harris et al. showed that community latrine coverage was a better predictor of childhood stunting than household ownership of a latrine. It is however not clear whether water coverage at community-level would also be a better predictor of child growth compared to household-level access to water. Hence, this research filled an important gap in the literature in trying to explain how individual and community-level access to water may affect childhood stunting in rural areas in resource-limited settings.

Purpose of the Study

The purpose of this research study was to explore the functional relationship between various measures of water access and childhood stunting in resource-limited settings. Water access, in this research study, was defined in terms of time or distance to water source, the quality of the water source, and the quantity of water used by households. I determined the associations between distance to water source, water quantity, and water quality with childhood stunting at 18 months of age in the two

rural districts of Zimbabwe in the Midlands Province. The research inquiry that I followed was a quantitative, analytical research paradigm using secondary data from the SHINE Study, a 2x2 factorial cluster-randomized clinical trial that was conducted in the two rural districts (The SHINE Team, 2015).

Research Questions and Hypotheses

This study was guided by three key research questions (RQs) that I used to explore the associations between childhood stunting and access to water at both child and community-level. I tested each of the three hypotheses corresponding to these research questions.

- RQ 1: Is there an association between household water quantity and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi?
    - $H_0$1*:* There is no association between household water quantity and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.
    - H_a1*:* There is an association between household water quantity and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.
- RQ 2: Is there an association between household distance to water source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi?
    - $H_0$2*:* There is no association between household distance/time to water source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.
    - H_a2*:* There is no association between household distance/time to water

source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.

- RQ 3: Is there an association between household water quality at source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi?
    - $H_0$3*:* There is no association between household water quality at source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.
    - H_a3*:* There is no association between household water quality at source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.

The SHINE Trial dataset provided an extensive list of covariates that were collected at baseline and could be used for adjusting the analyses for confounding. The detailed description of these is discussed in Section 3.

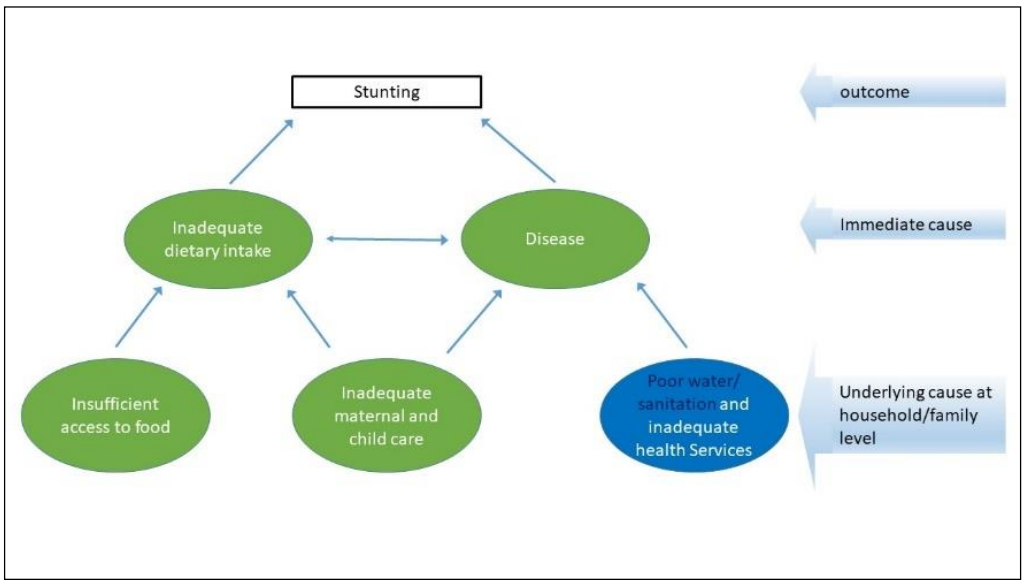
Theoretical Framework

The purpose of this study was to evaluate the effects of water access on childhood stunting because stunting is associated with poor access to water. Childhood stunting has been shown to be associated with adverse effects on other child health and developmental outcomes (de Onis & Branca, 2016). I proposed and tested research questions that are in the pathway to childhood stunting according to the theoretical model (Figure 1) based on the UNICEF malnutrition and stunting model (UNICEF& WHO, 1990). However, I only focused on the water and sanitation pathway as a determinant for childhood stunting as shown in Figure 2. The hypothesized theoretical model was based on the literature that links water quality,

distance to water, and water quantity to childhood stunting.

Figure 1

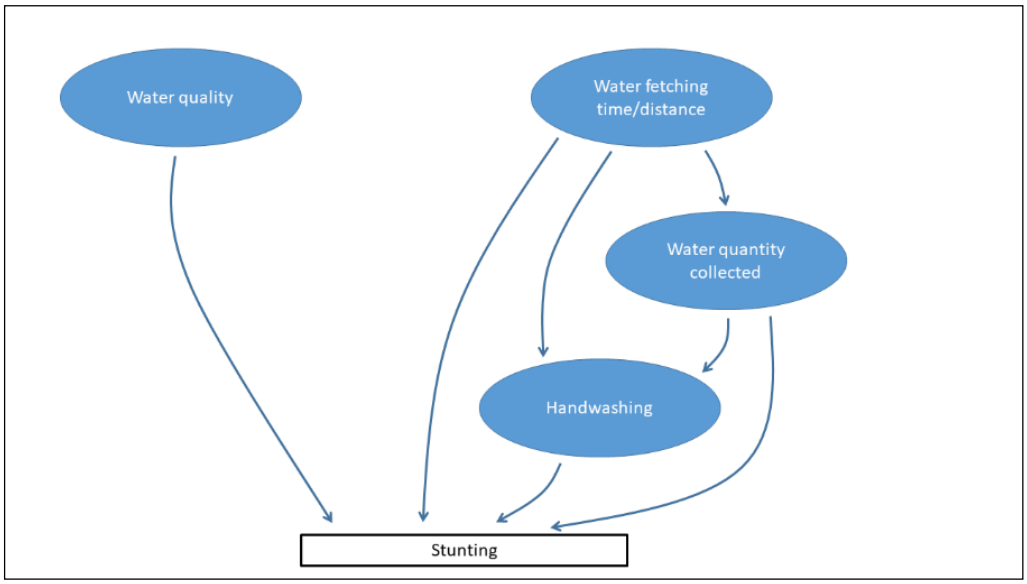
WHO Conceptual Framework on Childhood Stunting



Note. Adapted from "Strategy for improved nutrition of children and women in developing countries. A UNICEF Policy Review" by UNICEF, 1990. http://repository.forcedmigration.org/pdf/?pid=fmo:3066. Copyright 2022 by UNICEF. In the public domain.

Figure 2

Hypothesized Pathways to Stunting Attributable to Water Access



The overarching theoretical framework for this study was the social ecological model (SEM) that justifies the use of multilevel interventions to improve population health (Bronfenbrenner, 1994; Bronfenbrenner & Morris, 2006; McLeroy et al., 1988). Bronfenbrenner (1994) postulated that the environment influences human behaviors and actions and that humans create or modify the specific environments in which they live. SEMs have been successfully applied in public health to effect behavior change (Glanz et al., 2008). Childhood stunting is a complex health problem that requires a multilevel and multidisciplinary approach to interventions (de Onis & Branca, 2016; Stewart et al., 2013), making the SEM suited to study this situation.

SEMs have five defining principles that help influence health behaviors (Glanz et al., 2008). The first principle asserts that health behaviors are not simple; rather there are multiple levels of influence on health behaviors. These multiple levels of influence are (a) intrapersonal, (b) interpersonal, and (c) community levels affecting organizational behavior and influencing public policy (Barry & Honoré, 2009). In this study, I examined the association between water access at two levels, the individual (household) level and the community level. Results of this study could influence institutional behaviors and ultimately affect policy on childhood stunting.

The second principle of SEMs affirms that environmental contexts are important determinants of health behaviors, which means that peoples' health behaviors are determined by the environment in which they live (Glanz et al., 2008). For example, in environments where short stature is the norm, childhood stunting is not immediately recognized as a health problem. Additionally, in environments where water is hard to find people tend to use less water and engage in more risky hygiene behaviors such as not washing hands after using the toilet, which could increase the

risk of childhood stunting through contraction of diseases.

The third principle of SEMs points out that influences on behaviors interact across the various levels; hence, all levels should be addressed to influence the desired health behavior (Glanz et al., 2008). If at household level people are not aware of stunting as a challenge needing attention but the health institutions are ready to deal with it, then childhood stunting will remain a problem unaddressed. In addition, if no policies are put in place to prevent childhood stunting then public health workers will be less effective in their endeavors to prevent it.

The fourth principle of SEMs states that the models should be behavior-specific, which means that to influence a specific health behavior requires specifically tailored interventions (Glanz et al., 2008). For example, providing water chlorination chemicals will only improve treatment of drinking water in the household but not any other behaviors such as feeding nutritional foods to children. In this study, I examined the effects of water treatment on childhood stunting, where water treatment chemicals were supplied to the households.

The last principle of SEMs stresses that interventions that target multiple levels should be more effective in changing behaviors (Glanz et al., 2008). The WHO/UNICEF Framework for child malnutrition demonstrates this principle by separating the different domains that influence childhood stunting into proximal and distal determinants (Stewart et al., 2013). In this framework, access to nutrition, maternal childcare, access to water and sanitation and access to health services are identified as more proximal (or underlying) causes of childhood stunting. This framework also shows that disease and inadequate nutrient intake are the more proximal determinants of childhood stunting.

The SEM framework allowed me to investigate the mechanism through which water access may be associated with childhood stunting. This can help public health practitioners in resource-limited settings design interventions that are more effective to tackle the problem of childhood stunting in rural Zimbabwe and other SSA countries with similar settings.

Nature of the Study

This study was quantitative in nature; Accordingly, I tested objective hypotheses by examining the relationships between quantitative variables (Creswell, 2009). The purpose of this study was to evaluate the effects of access to water on childhood linear growth in a resource limited rural settings. Using specific RQs, I determined the association between child linear growth and various water access measures. A prospective observational cohort study design was employed to investigate the relationship between water access measures assessed at the baseline visit during pregnancy and child linear growth assessed at 18 months infant age. In this study I followed a question-driven approach where I first formulated the RQs then identified a suiTable Bataset to answer them, rather than a data-driven approach where a researcher interrogates available data to produce research questions (Cheng & Phillips, 2014). I identified a suitable source of secondary data to answer the research questions from a randomized clinical trial conducted in Zimbabwe between 2012 and 2017 in the two districts of Chirumanzu and Shurugwi by the Sanitation, Hygiene, Infant Nutrition Efficacy (SHINE) Study Team (SHINE Team, 2015).

The dependent variables were LAZ and childhood stunting and independent variables were distance and walk time to the primary water source, quantity of water, and the quality of water source used. Explanatory variables were selected from a list

of candidate variables, which include maternal and infant characteristics, household and environmental characteristics, and geographical characteristics as well as the randomized interventions given in the original trial and other study conduct related factors. Community-level variables collected at the cluster and district-levels were also used as explanatory variables in the analysis. The full description of methods used in this study is discussed in Section 2.

Details of the Secondary Data Source

The secondary data source identified for this study was the SHINE study conducted in Zimbabwe between 2012 and 2017. SHINE was a cluster-randomized, 2x2 factorial community based experimental study conducted in the two districts, Chirumanzu and Shurugwi in rural Zimbabwe (SHINE Trial Team, 2015). According to SHINE Trial Team, the objectives of the study were to investigate the independent and combined effects of an improved WaSH intervention and an improved nutrition intervention on child growth and anemia at 18 months of age. In this study the authors reported that, among HIV-unexposed children at 18 months of age, the nutrition intervention improved child linear growth by 0.16 HAZ (95% CI: 0.08–0.23), reduced stunting by 25% from 35% to 27%, increased hemoglobin concentration by 2.03 g/L (95% CI: 1.28–2.79), and reduced anemia from 13.9% to 10.5% (Humphrey et al., 2019). They also reported an even higher impact of the nutrition intervention among HIV-exposed children, increasing HAZ by 0.26 (95% CI 0.09–0.43) and hemoglobin concentration by 2.9 g/L (95% CI 0.90–4.90 (Prendergast et al., 2019). However, the authors found no evidence of impact of the WaSH intervention on either child growth or anemia among all the children. The SHINE study did not intervene in water access but only recorded how participants accessed water for various uses. Variables

collected in the SHINE study include water access variables in various dimensions, child growth outcomes; and many explanatory variables, which include maternal, child, household, and environmental characteristics suitable for analyzing this research study.

Literature Search Strategy

The literature search strategy involved breaking down the research questions into specific subject areas to be able to identify keywords and phrases that describe the subject being researched. I searched for journal articles, through the Walden Library, Cochrane Online Library, Google Scholar, PubMed, and other scholarly databases for publications related to child linear growth, childhood stunting, and water access in resource-limited settings. I also searched the UN, UNICEF and WHO websites for policy documents related to childhood stunting and water and sanitation. I particularly searched for original peer-reviewed research articles and review articles that synthesized evidence from several studies. I also followed references in relevant articles to identify seminal articles that they cited. I focused on articles that discussed childhood stunting in resource-limited settings, water access in resource-limited settings, and those that reported the association between childhood stunting and water access in resource-limited settings. The keywords that I used in the search included *child stunting, child malnutrition, linear growth, stunting prevalence, developing country, low income, middle income, sanitation access, sanitation coverage, water coverage, community coverage, water access, volume of water, water per capita, water quantity, water quality, water treatment,* and *social ecological.* The literature identified was published between the years 2007 to 2020, but older seminal research articles and books were also included for verification and reference. Following is a

synthesis of the available literature on childhood stunting and water access in resource-limited settings.

Theoretical Framework

The overarching theoretical framework for this research is the SEM. Childhood stunting is a multi-factorial problem that requires multi-sectorial solutions (Stewart et al., 2013), and the SEM provides the basis for exploring and explaining multilevel interventions that might have better impact on addressing the stunting challenge. For this research, I tested a theoretical model showing the pathway to childhood stunting mediated by access to water, which is based on the WHO/UNICEF (1990) stunting model. The theoretical model is based on literature, and links water access, water quality, water quantity and distance to water source to poor child growth outcomes and childhood stunting.

Description of Social Ecological Model

According to the original ecological framework for human development proposed by Bronfenbrenner (1989), SEMs have five levels of influence: (a) individual, (b) microsystem, (c) mesosystem, (d) exosystem, and (e) macrosystem. These were further adapted for application in public health to understand what influences health behaviors (Glanz et al., 2008). It is however noted that critics of the SEM as a basis for explaining or effecting behavior change cite the difficulties of applying the model due to its breath (Glanz et al., 2008). In this study, I examined the association between water access and child health at only two levels, the individual (household) and community levels. Results of this study could however influence other levels such as institutional behaviors and ultimately affect policy on child malnutrition, stunting, and water access in LMIC.

Constructs of Social Ecological Model

SEMs are multi-layered constructs with concentric layers of influence. Various researchers have labelled these layers of influence differently but they all center around the individual. According to Glanz et al. (2008) SEMs have five defining principles that help influence health behaviors: (a) individual, (b) interpersonal, (c) community, (d) organizational, and (e) policy. Barry and Honoré (2009) also specified these multiple levels of influence as (a) intrapersonal, (b) interpersonal, and (c) community levels, which affect (e) organizational behavior and influence (f) public policy. In Figure 3 I illustrate the levels of the SEM as applied by the Centers for Disease Control and Prevention (CDC, 2018) in their violence prevention health promotion model. I present and explain each level of the SEM in Table 1.

Figure 3

The Social Ecological Model

Note. From "The Social Ecological Model: A Framework for Prevention" by Centers for Disease Control and Prevention (CDC), 2022, *Violence Prevention* (http://www.cdc.gov/violenceprevention/overview/social-ecologicalmodel.html). Copyright 2022 by Centers for Disease Control and Prevention. In the public domain.

Table 1

Social Ecological Model Levels

SEM Level	Description
Individual	Characteristics of an individual that influence behavior change, including knowledge, attitudes, behavior, self-efficacy, developmental history, gender, age, religion, racial/ethnic identity, socio-economic status, access to financial resources, values, goals, expectations, literacy, and others.
Interpersonal	Both formal and informal social networks and social support systems that can influence individual behaviors. These include family members, friends, peers, co-workers, religious networks, customs or traditions.
Community	Relationships among organizations, institutions (for example hospitals, clinics, schools, and others), and informational networks within defined boundaries, including the built environment (business centers, growth points, and others), village associations, community leaders, businesses, and transportation.
Organizational	Organizations or social institutions (for example non-governmental organizations (NGOs), mission hospitals, and others) with rules and regulations for operations that affect how, or how well, for example, maternal and child health services are provided to an individual or accessed by groups.
Policy/Enabling Environment	Local, district, provincial, national and global laws and policies, including policies regarding the allocation of resources for maternal, newborn, and child health and access to healthcare services, restrictive policies (e.g., high fees or taxes for health services), or lack of policies that require improved access to water and sanitation services.

Note. Adapted from "The UNICEF model of Communication for Development (C4D)" by UNICEF, 2009. Copyright 2022 by Academic Publishing Consortium. In the public domain.

Studies that Applied Social Ecological Models in Public Health

SEMs are useful theoretical frameworks for promoting, predicting, and explaining health behavior change in individuals and populations. Researchers have utilized SEMs to design interventions that target determinants of health at multiple levels and successfully evaluated these interventions for effectiveness. I reviewed literature on programs that used SEMs to promote or explain health behavior change using multiple levels of influence. Several researchers used SEMs to design interventions that target multiple levels of health determinants. Baruth and Wilcox (2013) implemented an intervention guided by the structural ecological model focusing on two levels of the SEM, environmental and organizational levels, to effect changes in health behavior among church members. The behaviors targeted were

improvements in physical activity, consumption of fruit and vegetable, and intake of fat and fiber. They reported that nearly half of the participants changed two or more of these targeted behaviors. This study showed that improvements in multiple behaviors simultaneously was achievable and that faith-based interventions targeting environmental and organizational change can successfully change multiple behaviors, potentially leading to greater improvements in public health.

Stark et al. (2017) evaluated the effectiveness of an intervention that applied the SEM approach to an online training program on childhood obesity prevention. The aims of the study were to measure the extent to which an ecological approach was applied at the community level by health staff after a 6-month period post taking the training course. Stark et al. also sought to evaluate the individual and organizational characteristics associated with a participant's likelihood of applying an ecological approach to prevention of childhood obesity. Findings of this study demonstrated successful application of the SEM where three organizational characteristics were positively associated with the application of an ecological approach, and individual characteristics were negatively associated with the application of an ecological approach.

Kumar et al. (2012) utilized a social ecological framework to examine influenza vaccine uptake during the 2009 H1N1 pandemic. Kumar et al. reported that uptake of the influenza vaccine was predicted by variables at each level of the SEM. Their results showed that the intrapersonal, interpersonal, institutional, and the policy and community levels each explained 53%, 47%, 34%, and 8% of the variance associated with vaccine uptake respectively. They then concluded that overall, the various levels of the SEM altogether explained 65% of the variance, suggesting that

interventions targeting multiple levels of the framework would be more effective than interventions aimed at any single level. From all these studies it is clearly demonstrated that the SEM approach can be a useful framework for designing and evaluating public health studies.

In large-scale population health promotion programs, the SEM framework has also been utilized to design public health behavior change interventions targeting the population at large. The CDC (2015) adapted the SEM of health promotion in their Colorectal Cancer Control Program. This program uses a multi-level approach to colorectal cancer prevention with the intervention levels mapped onto the SEM levels of influence. At the center of the model is the individual, surrounded by four bands of influence representing the interpersonal, community, organizational, and policy levels. In another program the CDC used a SEM approach in a violence-prevention intervention strategy with a goal of stopping violence before it began by understanding the factors that influence violence (CDC, 2018). They employed a four-level SEM where the complex interplay between individual, relationship, community, and societal factors were considered. This approach enabled CDC to understand how violence develops, and the effects that potential prevention strategies would have.

On a similar scale, the American College Health Association (ACHA) implemented the Healthy Campus 2020 initiative to improve student and staff health on campuses (ACHA, 2018). The ACHA sought to provide answers to the question: "How to give everyone on campus a chance to lead a healthy life and to live longer?" by emphasizing an ecological approach. The ecological approach taken by ACHA focused on both population-level and individual-level determinants of health and

interventions. The initiative recognized that campus life is complex, and the ecology provides a multifaceted view of the connections among health, learning, productivity, and campus structure.

The U.S. Department of Health and Human Services continues to implement the Healthy People 2020 Framework utilizing a SEM approach to improve population health (Department of Health and Human Services, n.d.). The Department of Health and Human Services emphasizes the importance of an ecological and determinants approach to health promotion and disease prevention. The objective of the Healthy People 2020 Framework is to enable the population to attain high-quality, longer lives free of preventable diseases, disability, injury, and premature death.

In all these programs the targeted groups are complex populations with multiple health determinants. Both CDC (2015, 2018) and ACHA (2018) asserted their beliefs that the SEM approach optimizes synergies of the interventions and achieves maximum impact when public health activities are implemented at these five levels. The ACHA concluded that the SEM approach has the most effective interventions targeting determinants at the multiple levels. The CDC (2015) also reported that the SEM approach targeting multiple levels improves the likelihood to sustain prevention efforts over time than any single intervention. The diversity of studies and programs addressing complex public health problems by utilizing the SEM framework highlights the suitability of the theoretical framework for the current study of childhood stunting. Childhood stunting has determinants at multiple levels and requires complex analysis and implementation of interventions that target these multiple levels.

Literature Review

Childhood Stunting

Globally, about 22.2% or 151 million children < 5 years were affected by stunting in 2017, and 58.7 million of them lived in Africa and 83.6 million lived in Asia (UNICEF et al., 2018). Between 36-42% of these stunted children lived in SSA and Asia in 2013 (Black et al., 2013). The problem of childhood stunting is well characterized with decades of many national surveys and regional and global meta-analyses (de Onis & Branca, 2016). It has been shown that stunting begins in-utero, continues through the first two years of a child's life then becomes almost irreversible thereafter. This period of the first 1,000 days in a child's live is usually referred to as "the window of opportunity" for stunting interventions, meaning that if proper interventions were implemented during this period, then childhood stunting could be reversed (de Onis & Branca, 2016).

Childhood Stunting and Sustainable Development Goals

Two decades ago the problem of childhood stunting was recognized as an important public health problem that requires urgent solutions and was elevated for discussion at the highest levels. Governments, through the UN, made commitments to reduce childhood stunting and other health problems by setting agreed targets of reduction in stunting prevalence worldwide. In the year 2000, MDGs where declared, and among these was target 1c with the aim to reduce malnutrition to half by the year 2015 (UN, n.d.-a). A review of progress toward these targets was done in 2015, and a new set of aspirational targets, the SDGs, were adopted. To address the problem of childhood stunting, SDG 2 was made with the aim to reduce malnutrition including stunting by 2030 to internationally agreed targets (UN, n.d.-b). Periodic evaluations of

each nation's progress toward these targets are regularly compiled and compared.

Trends in Childhood Stunting

As a result of the sustained efforts to reduce childhood stunting, globally the prevalence of stunting has been declining over the last decade (de Onis & Branca, 2016; UNICEF et al., 2018). The proportion of stunted children declined from 32.6% in 2000 to 22.2% by 2017, and this corresponds to a decline from 198.4 million stunted children in 2000 to 150.8 million by 2017 (UNICEF et al 2018). In SSA, the prevalence of stunting has also been declining in line with global trends, from 38.3% in 2000 to 30.3% by 2017. However, the absolute population of stunted children living in SSA increased from 50.6 million in 2000 to 58.7 million in 2017 due to population growth in this young population (de Onis & Branca, 2016; UNICEF et al, 2018). Stunting remains a hidden public health problem often discussed only in academic circles but not by public health practitioners or the affected communities (Said-Mohamed et al., 2015). In Zimbabwe, according to the last Zimbabwe Demographic and Health Survey: 2015 conducted by the Zimbabwe National Statistics Agency (ZIMSTAT) and ICF International, stunting has also declined in the country from 35% in 2005 to 27% by 2015 (ZIMSTAT & ICF International, 2016). However, mapped trends in stunting globally showed reductions in childhood stunting worldwide except in SSA (Zimmerman et al., 2018). The evidence shows that although stunting may be declining in the rest of the world, SSA still has to grapple with the problem and find effective interventions.

Underlying Causes and Complexity of Childhood Stunting

Childhood stunting is a malnutrition condition that is intricately linked with poverty, disease, and deprivation. When children are stunted it often reflects on a

multitude of other social, economic and health related problems that they live in or are exposed to during their growing up. Childhood stunting is the best indicator of the overall wellbeing of a child (de Onis & Branca, 2016). Stunting may indicate the cumulative effects of insults to a child's life that include nutrient deprivation, hunger, disease, and exposure to unsanitary environmental conditions, culminating in stunted child growth (de Onis & Branca, 2016). Undernutrition is by far the biggest risk factor for stunting, wasting and underweight in children (Black et al., 2008). Stunting is associated with poor nutrition, repeated infections, and inadequate psychological stimulation (Stewart et al., 2013). Stunting therefore is an important public health indicator of insults suffered by a child during early infancy.

Researchers proposed a number of frameworks and models to explain the etiology and determinants of childhood stunting. In 1990, UNICEF proposed a framework for explaining underlying determinants of child undernutrition and stunting. The basic underlying causes of undernutrition included in this framework were environmental, economic, and socio-political factors (UNICEF & WHO, 1990). This framework was very important in exploring the determinants of childhood stunting as it brings together the linkages among a range of factors and shows how they may interact to influence growth and development of a child. The framework also shows that some determinants of stunting may be more proximal to the problem while others are more distal. In this framework, proximal determinants include inadequate dietary intake, which includes exclusive breast-feeding and others infant and young child feeding practices, and infection and disease. Distal determinants of stunting on the other hand include insufficient access to food, inadequate maternal and childcare, and the environment in which the child grows up, that includes access

to water and sanitation and inadequate access to healthcare services. Childhood stunting is therefore a complex public health problem that has multiple underlying causes. To understand it and develop effective interventions to tackle the problem requires of childhood stunting requires a multi-sectorial approach.

Consequences of Childhood Stunting

The adverse effects of childhood stunting are not only limited to shortness of stature, but stunting is also associated with multiple health problems that affect the lives of stunted children during infancy, persist into their adulthood and affects even their offspring (Black et al., 2008; de Onis & Branca, 2016; Dewey & Begum, 2011; Stein et al., 2005). Researchers have shown that stunting can cause both short-term or immediate and longer-term health, social and economic problems as outlined in the following sections.

Short-term or immediate consequences of stunting include susceptibility to morbidity and mortality; stunted children have higher mortality and morbidity compared to non-stunted children. It is estimated that 35% of the global burden of disease in children < 5 years in 2008 was attributable to maternal and child undernutrition (Black et al., 2008). Globally 3.5 million children < 5 years died from maternal and child undernutrition in 2013, and an estimated 14.5% of these child deaths were stunted (Black et al., 2008). It has been shown that stunted children also have higher morbidity compared to non-stunted children. In the short-term, child stunting is associated with increased morbidity from infectious diseases such as pneumonia and diarrhea (de Onis & Branca, 2016; Prendergast & Humphrey, 2014; and Stewart et al., 2013). Stewart et al. also pointed out that infectious diseases cause more stunting, which in turn increases susceptibility to other diseases resulting in a

downward spiral in the health status of these affected children.

In the long-term, some of the consequences of childhood stunting include increased risk of being stunted as adults. Poor fetal growth or stunting in the first 2 years of life leads to irreversible damage in later life. Children who were stunted in early life are associated with shorter adult height in their adulthood (Cesar et al., 2008; Dewey & Begum, 2011). Maternal stunting is a risk factor for infant health and survival. Stunted women are associated with many adverse birth outcomes including chronic fetal distress, stillbirths, and higher neo-natal mortality (Black et al., 2008; Dewey & Begum, 2011). Stunted women are also associated with bearing low birth weight and stunted children (Black et al., 2008; Cesar et al., 2008; de Onis & Branca, 2016; Dewey & Begum, 2011).

Neurocognitive development is slower in stunted children. Stunting is associated with poor early childhood cognitive development outcomes in several domains on child development that includes motor stills, language and social (Adair et al., 2013; Black et al., 2008; Cesar et al., 2008; Dewey & Begum, 2011; Stein et al., 2005; Vollmer et al., 2014). Children who were stunted during infancy tend to attend less schooling and are likely to perform poorly in class (Adair et al., 2013; Maluccio et al., 2009). In their adult life, children who were stunted tend to work in less paying jobs and have reduced economic productivity (Adair et al., 2013; Dewey & Begum, 2011; Maluccio et al., 2009). This therefore means that because of stunted growth in infancy there is overall reduced human capital development.

Finally, in their adult life, children who were stunted in infancy are likely to become more susceptible to non-communicable diseases such as hypertension and heart diseases (Adair et al., 2013; Victora et al., 2008). Prendergast and Humphrey

(2014) discussed the stunting syndrome and presented a summary schematic diagram that depicts the intergenerational cycle of stunting highlighting the critical stages for interventions to break this vicious cycle of stunting.

Childhood stunting and WaSH

Access to clean water and improved sanitation are generally known to improve human health and reduce the risk of infection from diseases. Clean drinking water, in addition to hydration, reduces the risk of infection from water borne diseases; improved sanitation enables safe disposal of excreta creating a safe uncontaminated environment; handwashing reduces the risk of infection after contact with infectious agents. In the UNICEF Framework for undernutrition, one pathway to childhood stunting is through water, sanitation, and hygiene, which is mediated through diseases (UNICEF & WHO, 1990). Many observational studies have repeatedly shown strong associations between WaSH and disease as well as WaSH and child growth, including childhood stunting (Esrey et al., 1991; Fry et al., 2010; see also Lehmann et al, June 2016). In a study conducted in Mali to investigate community sanitation coverage and child growth, a positive association between latrine community coverage and childhood stunting was reported (Harris et al., 2017). It is however important for both researchers and public health practitioners to understand the role that each component of the WaSH interventions plays in improving health outcomes.

There are very few randomized intervention trials of WaSH with childhood stunting as an outcome. Recently three randomized studies of WaSH interventions and child growth outcomes reported no association between WaSH and childhood stunting (Humphrey et al., 2019; Luby et al., 2018; and Null et al, 2018). In two of these trials the WaSH interventions were split into their WaSH components of –

sanitation, handwashing, and water treatment (Luby et al., 2018; Null et al., 2018). In the other trial all WaSH components were delivered as a combined intervention (Humphrey et al., 2019). Findings from these studies contradicted evidence from the many observational studies that reported positive associations between childhood stunting and access to improved WaSH. However, in all three randomized studies, the researchers did not randomize access to water and the role it may play in childhood stunting was not evaluated.

Water Coverage and Childhood Stunting

A number of studies showed that a large part of child malnutrition, including stunting, is associated with poor access to water and having inadequate sanitation and hygiene facilities. Lehmann et al. (June 2016) compiled a report in which countries were ranked on childhood stunting prevalence, proportion of the population without water access, and proportion of the population without access to toilets. Countries with the highest prevalence of childhood stunting globally have correspondingly poor access to water as well as poor access to toilets (Lehmann et al., June 2016). They however noted that water access or water coverage was defined differently in different settings and by different agencies, hence may not mean the same in different settings. In a report in 2003, Aiga and Umenai highlighted that there was no standard and universally agreed definition of water access, and this remains true up to today. It was also noted that different government and development agencies used different ways to define access to water and progress to meet MDGs and SDGs was measured differently according to each country (Aiga & Umenai, 2003). Furthermore, distance, time, and water quantity were all used in various ways to define access to water (Aiga & Umenai, 2003). For investigating the association between water access and

childhood stunting, different researchers have used some of these definitions in their analyses. Following is a review of some studies that investigated access to water and child health outcomes including childhood stunting.

Childhood Stunting and Distance or Time to Water Source

In rural communities in SSA, the distance walked to a water source and time taken to fetch water are strongly correlated, hence distance walked to a water source is sometimes used interchangeably with walk-time in water and sanitation surveys (Esrey et al., 1991; Nygren et al., 2016). The quantity of water used by a household is also correlated with the ease of accessing the water source (Nygren, et al., 2016). Hence, households that are located further from water sources are expected to use less water than those located near the water source. In 1991 Esrey et al. carried out a comprehensive review of studies that investigated the effects of WaSH access on morbidity, mortality, and child growth. They reported mixed findings on child growth but pointed out that in most studies they could not distinguish the effect of water quantity or water quality on the child health outcome of interest. In some of these studies reviewed, distance to water source was used as a proxy for the quantity of water used.

In 2012 an analysis of data on water access from 200,000 data points from Demographic and Health Surveys was carried out in many SSA (Pickering & Davies, 2012). They investigated the association between water fetching time and child health outcomes. The authors found that time spent walking to fetch water was a significant predictor of child health outcomes. Availability of freshwater and distance travelled to fetch freshwater adversely affected child health outcomes including childhood stunting (Pickering & Davies, 2012). They reported that particularly distance walked

to fetch freshwater uniquely predicted childhood stunting and underweight in their analyses. Pickering and Davies then concluded that a 15-minute reduction in one-way walk was associated with an increase of 0.3 standard deviation in HAZ, which is a significant reduction in childhood stunting.

In a study conducted in the Kilimanjaro Region of Tanzania, 44.2% of the children aged between 1 to 35 months were stunted (Abubakar et al., (2012). Distance to water source was an independent predictor of underweight but not for stunting. In this study, despite the high levels of stunting observed – the authors found no evidence that availability of water reduced childhood stunting. They further recommended that understanding the context matters when designing and implementing interventions to reduce child malnutrition and stunting.

Evidence suggesting an association between childhood stunting and distance and time travelled to fetch freshwater in rural settings is mixed; however, there are no experimental studies that have randomized access to water with child health outcomes. This view was also supported in another review of literature on the mechanism through which WaSH interventions could reduce childhood stunting (Cumming & Cairncross, 2016). Cumming and Cairncross acknowledged that even with all the observational evidence suggesting that WaSH could reduce stunting, there has never been a trial to intervene on water access and that more evidence was required to make causal inference.

Childhood Stunting and Quantity of Water Used

Water is essential for child survival and development; however, the quantity water required to achieve good health is not well documented. In many studies investigating child health outcomes and WaSH, researchers looked at the quantity

water required to reduce morbidity and mortality - especially diarrhea as the main outcome of interest. In a systematic review of literature on the amount of water per person per day needed to prevent morbidity and mortality in emergency settings, the authors reported positive correlations between the quantity of water used and reported cases of diarrhea from different studies (De Buck et al., 2015). They however reported that different amounts of water were required to make an impact in each setting. For example, in Ghana a difference of 11 liters per person per day between households with and without cases of diarrhea was needed to see a difference. In Kenya, however the authors reported that a difference of ~6 liters per person per day was needed to see a difference in reported diarrhea cases. The study however reviewed studies conducted in refugee camps and not in normal rural settings. A similar approach may be necessary to understand the quantity of water required per person per day to reduce childhood stunting.

In a review of studies that implemented WaSH interventions with childhood stunting as the outcome, mixed findings were reported for both morbidity and growth (Cumming & Cairncross, 2016). The authors of the review noted that no studies that they reviewed had evaluated the water quantity component of the WaSH interventions to test its impact on stunting. Both Cumming and Cairncross as well as De Buck and colleagues concluded that primary evidence was required to make causal inference on water quantity required for morbidity, mortality, and growth (Cumming & Cairncross, 2016; De Buck et al., 2015).

Childhood Stunting and Water Quality

Clean water is essential to prevent infections from water borne diseases. Rural water is often not treated with chemicals and its quality is usually only determined by

the quality of the source used. Even for the purposes of meeting the MDGs and later SDGs, countries classify rural populations with access to safe water as those fetching drinking water from protected wells and boreholes (WHO, 7 February 2018). In a Cochrane systematic review of studies that investigated the effect of interventions to improve water supply and quality, provide sanitation, and promote handwashing with soap on physical growth in children, Dangour et al. (2013) concluded that improving the quality of water had small benefits to linear growth for children < 5 years. However, they also reported that the quality of the evidence in the studies that they reviewed was poor. This highlights a gap whereby more and better-quality evidence is required to quantify the health benefits of improving water quality in resource-limited settings in order to improve child linear growth and reduce stunting.

Current Interventions to Prevent Childhood Stunting

A systematic review and meta-analysis of complementary feeding intervention studies on child growth in LMIC showed that nutrition interventions can reduce stunting by up to a third of the deficit, but the remaining gap remains unexplained (Panjwani & Heidkamp, 2017). They showed that the best nutrition interventions could only reduce stunting by between 10-22% in food-secure and insecure environments respectively. WaSH interventions had been suggested as a possible solution to reduce the stunting gap (Humphrey, 2009), but little experimental evidence was available to support this. Evidence from three recently completed randomized controlled intervention trials of WaSH with childhood stunting outcomes reported that nutrition interventions improved child growth and reduced stunting by between 16 to 22% (Humphrey et al., 2019; Luby et al., 2018; Null et al, 2018). These three trials were conducted in different settings, two in Africa and one in Asia, and all

three trials did not find an impact of the WaSH interventions on childhood stunting. However, all three studies did not intervene on water access, hence more evidence is required to understand the role of water on childhood stunting.

Operational Definitions

For the purposes of examining the association between childhood linear growth and water access, the following key terms are defined as follows:

Child linear growth: is defined as length-for-age z-score (LAZ) or height-for-age z-score (HAZ) in comparison to reference growth standards – in this case the WHO Multicentre Growth Reference Study Group (2006).

Childhood stunting: is defined as having LAZ < -2 standard deviations of the WHO length-for-age reference curve (WHO Multicentre Growth Reference Study Group, 2006). For children below the age of 24 months recumbent length is measured instead of standing height to calculate the z-score.

Cluster: refers to a geographic area within the study districts covered by 1-3 community health workers that was used as a unit of randomization in the SHINE Trial (SHINE Trial Team, 2015). During the SHINE Trial 212 clusters were defined across the two study districts.

Household water access: generally refers to having portable water for the household within a reasonable distance, water from a protected source, and water that is safe from pathogens and chemical impurities (WHO, 7 February 2018). The WHO has used several definitions of water access in different environments: in the cities it is defined as access to a safely managed drinking water service that is located on premises, available when needed, and free from contamination. In rural areas it is defined as access to an improved drinking-water source within a round trip of 30

minutes to collect water. In this study the rural definition will be used to define water access.

Household distance/walk-time to water source: distance to a water source in this context is defined as the round-trip total distance (or time) walked from the household to a water source and back. This is either collected as self-reported distance or time, or estimated as the Euclidean distance between two points, or by GPS (Pearson, 2016).

Household water quantity: refers to the total volume of water used by each individual household member in a 24-hour period, volume per capita per day (Tamason et al., 2016). This is estimated as the total volume of water collected by a household divided by the size of the household and number of days the water is used. Tamason et al. however noted that there is no reliable standard measure of water quantity without metering.

Household water quality: generally, refers to the chemical, physical, biological, and radiological characteristics of water source used by the household. It is a measure of the condition of water relative to the requirements of one or more biotic species and or to any human need or purpose (Diersing & Nancy, 2009). In this study water quality is defined as water collected from a protected water source.

Protected water source: protected sources are those with barriers or other structure to protect the water from contamination. All surface water sources, such as lakes, rivers and streams or poorly constructed wells, are examples of unprotected sources. Piped water systems, boreholes, and lined deep wells are examples of protected sources. (Bruni & Spuhler, 2020).

Community-level childhood stunting: is defined as the proportion of stunted

children (LAZ < -2) within a geographic area or cluster at the age of 18 months.

Community-level distance to water: is defined as the average of the household distance to water sources among all households within a given cluster.

Community-level water quantity: is the average of the household water quantity among all households within a given cluster.

Community-level water quality: is defined as the proportion of households within each cluster fetching household water from a protected water source.

Assumptions

In this study I assumed that the two study districts from which participants in the original study were recruited have similar settings to other rural districts in the country and represent a typical rural setting in a LMIC. I assumed that children who were participants in the study are representative of other children in the two districts to make general inference about children in low-income rural settings. My other assumptions were related to using a secondary data source to conduct research; I assumed that respondents provided truthful answers to the survey questions during data collection interviews. I also assumed that the coding of the responses in the provided dataset are accurate, and that the documentation is complete. One major disadvantage of using secondary data to conduct research is that some of the key details that would make the data better interpretable may have been removed as a way of de-identifying the dataset for public access (Cheng & Phillips, 2014). I assumed that when households live closer to a water source or they take less time to fetch water, then they will use more water for consumption and hygienic purposes such as hand washing, resulting in better health outcomes including child linear growth. I also assumed that cleaner water results in less water borne infections leading to better

health outcomes and child growth. I then assume that when a community has higher water access they will be healthier and have better child growth.

Scope and Delimitations

In this study I set out to understand the association between access to water by households and linear growth among their child at 18 months of age for children residing in the two rural study districts of Chirumanzu and Shurugwi in the Midlands Province of Zimbabwe. Water access was operationalized as three metrics measuring quantity, quality, safety, and accessibility: (i) distance to water source, (ii) quantity of water used, and (iii) quality of the water source used. These three metrics were measured at household-level and computed at cluster-level. Child linear growth was measured as LAZ and stunting was computed. This study was limited to children who were participants in the SHINE Trial that was conducted between November 2012 and July 2017 in those two districts (SHINE Team, 2015).

Limitations of the Study

This study was observational in nature hence any observed associations may not be interpreted as being causal relationships. The study concentrated on the observed patterns of water access and usage and measured child linear growth and does not attempt to explain why those patterns may have been observed. I hypothesized that improved water access in a community would improve the linear growth in children and used regression analysis to estimate the population average change in LAZ and stunting.

Another limitation of this study was the use of secondary data. Secondary data collected by others for their own specific purposes (Glass, 1976; Smith et al., 2011). There are limitations associated with using these data to conduct new research (Cheng

& Phillips, 2014), which include:

- Data may be incomplete; some key variables could be missing in the accessible dataset to enable a more complete and proper interpretation of results obtained.
- The data may be too old if data were collected and the situation has since changed so much that data may be less representative of the current situation in the study area.
- Some variables may have been measured on a different scale than what would be idea to answer the current research questions.
- Variable definitions could have changed over time during data collection in the trial.
- The study sample may not be representative of the larger population from which they were collected such that results will not be generalizable to that population.
- The dataset may have other biases reflecting the views of the researchers who conducted the primary study.

Significance of the Study

Childhood stunting is a huge public health problem affecting about 149.2 million children worldwide and if not resolved stunting affects the growth and development of children resulting in huge human capital loses when they become adults. This study is unique in that it addresses an important but under-researched area of public health looking at the role of water access on childhood stunting in resource constrained settings. Results of this study could provide evidence showing the role that each dimension of water access plays in the development of childhood stunting in

these settings. Results of this study could guide how interventions improve child linear growth could incorporate water access for better impact in this setting.

The study can add to the body of knowledge by quantifying the health benefits of access to water on childhood stunting in resource-limited settings. Existing evidence for example does not give clear guidance as to how much water is required to impact childhood stunting in SSA.

The findings of this study could inform the rural communities on how optimal access to water resources could improve child health and prevent childhood stunting. The findings could inform local authorities on how they should provide for water resources to rural communities. The findings are important to public health providers and policy makers as evidence to help them in the planning of malnutrition and disease prevention programs through the provision of optimal water resources to populations in low-income rural settings.

Implications for Social Change

With the current trend towards precision public health, findings of this study could change how water services are optimally provided to rural communities in resource-limited settings in order to impact childhood stunting. The findings could also impact how public health policy on malnutrition and disease prevention are implemented. These potential impacts could have a positive social change effect on communities in countries where childhood stunting is a significant public health problem and water provision is a challenge, which is true for most countries in SSA.

Summary

The role of WaSH in human health has been known since the last century; John Snow in 1855 attributed the spread of cholera to poor water quality and

unimproved sanitation in his ground breaking epidemiologic studies. WaSH continues to be as relevant to public health even to this day, and we continue to learn new functions that water plays in human health. The control of infectious diseases such as cholera, access to clean water is vital. However, understanding the function of water in the etiology of childhood stunting remains an area under study as the literature reviewed demonstrated. From this review, I concluded that stunting is an intransigent problem that is still a major public health challenge in SSA and needs new solutions to tackle it. The role of water in the prevention of childhood stunting in LMIC still needs to be explored further in order to understand the critical pathways to stunting. This section includes a review of studies, models, case studies and other scholarly works that discuss childhood stunting, water access, and their association. The section therefore, provides a linkage between the research questions and hypotheses identified in the last section with the available body of evidence in scientific literature and other credible sources. In section 3 I will describe in detail the research design and methodology used in conducting the study.

Section 2: Research Design and Secondary Data Source

There are few studies that examined the effect of water access on child linear growth in SSA, and no such studies have been carried out in Zimbabwe. Studies of predictors of childhood stunting conducted in different settings were investigated; however, in the current study population little research has been conducted to understand these drivers of childhood stunting and the role that access to water may have. Previous studies of childhood stunting in LMICs that focused on WaSH interventions reported mixed findings on child growth outcomes, but few of the studies attempted to quantify the child health benefits of access to water in rural settings in SSA (Esrey et al., 1991; Fry et al., 2010; Rasella, 2013). In this research study, I explored the association between child linear growth and stunting with various measures of access to water in two rural Zimbabwean districts using secondary data from a prospective cohort.

In this section, I present the methodology used to accomplish the purpose of this study. I then follow by discussing the research design employed in this study to answer the research questions. I present a detailed description of the study population and its settings, the selection of study participants into the study, and describe the secondary data source used. I also present and discuss the data analysis plan that includes the statistical procedures for testing the hypotheses, presentation, and interpretations of findings. Next, I will identify and discuss potential threats to internal, external and construct validity of the study, and description of the actions taken to address these limitations. Finally, I present a description of the procedures undertaken to ensure the conduct of an ethical study.

Research Design and Rationale

The study design was a quantitative, analytic study where I tested objective hypotheses by examining the relationships between quantitative variables (see Creswell, 2009). The purpose of the research study and the research questions were to determine the association between childhood stunting and water access in a resource limited rural setting. In this study, I sought to address a gap in the literature on the amount of water required to impact childhood stunting; the quality of drinking water required to impact childhood stunting; and the distance or time to a water source beyond which childhood stunting would not be impacted, or to establish if there was such a threshold. I also set out to test if community water coverage was a better predictor of child linear growth than household level access to water. The study followed a question-driven approach where research questions were formulated first then suitable a dataset identified, rather than a data-driven approach where a researcher interrogates the dataset to come up with research questions (Cheng & Phillips, 2014; Smith et al., 2011). However, this process is often not rigid but iterative and final research questions may be modified to best fit available data.

A randomized placebo controlled double-blinded clinical trial is recognized as the gold standard for investigating epidemiological evidence of causality (Hulley et al., 2007; Misra, 2012). It is not always possible to use such a design in some cases due to practical limitations and ethical considerations. Access to water is one such exposure that cannot be masked during a clinical trial. There have not been any randomized clinical trials of access to water where water access is randomly allocated to study participants. A researcher cannot limit access to water to a population for the purposes of conducting an experiment; hence, only quasi-experiments where access to

water is only observed during the conduct of a clinical trial can be used. Quasi-experiments are studies where the researcher conducts an experiment but cannot randomly assign the exposure to the participants (Creswell, 2009; Frankfort-Nachmias et al., 2015). This study used a prospective cohort study design where children were assessed for stunting at the age of 18 months and water access assessed during baseline and follow-up was used as the exposure. It has been shown that community-level factors may be more predictive of childhood stunting than household-level factors (Harris et al., 2017). In this research study I also tested the hypotheses at both household and community levels.

Methodology

Study Settings

The SHINE study was conducted on a population drawn from two rural districts of Chirumanzu and Shurugwi in the Midlands Province of Zimbabwe. According to the ZIMSTAT Census Report 2012 (ZIMSTAT, 2013a), the population of Zimbabwe is young; 41.1% of the population is under the age of 15 years and 15.1% are under the age of 5 years. In the Midlands Province where the two study districts are located, 44.6% of the population is under the age of 15 years and 15.3% are under the age of 5 years (ZIMSTAT, 2013b). In the Zimbabwe Demographic and Health Survey 2015 final report, the prevalence of childhood stunting in rural Zimbabwe was 28.7% in 2015 and in the Midlands Province was 27.4% (ZIMSTAT & ICF International, 2016). HIV prevalence is ~15% among antenatal women in this population. Previously across both Chirumanzu and Shurugwi districts, in 2011 only 16.5% of the study population had access to portable and perennial water less than 100m from their homes (Ntozini et al., 2015). There are also wide seasonal variations

in water access with a decline in functional water sources of 45% between the rain and dry seasons. Hence, the study setting is characterized by a young growing population with high prevalence of childhood stunting; high prevalence of HIV among antenatal women, which means many HIV-exposed children; and poor access to water.

Study Population

From the general population of all children under the age of 2 years living in the two study districts of Chirumanzu and Shurugwi, the target population for this research was all children under the age of 2 years residing in the two districts between the years 2012 and 2017. This population was selected because the clinical trial on which this secondary data analysis research study is based on was conducted in those settings by the SHINE trial team. An accessible population includes members of the target population remaining after omitting out those who are not willing or available to participate (Asiamah et al., 2017). The accessible population for this study is all children whose mothers consented to participate in the SHINE Trial study.

Design of the SHINE Study

Details of the SHINE Trial Study were described in the 2016 trial design series of papers that give detailed background to the SHINE trial interventions. Briefly, SHINE was a cluster-randomized, 2x2 factorial community based experimental study conducted between 2012 and 2017 in rural Zimbabwe (SHINE Trial Team, 2015). The objectives of the study were to investigate the effects of an improved WaSH intervention and an improved nutrition intervention on child growth and anemia at 18 months of age (SHINE Trial Team, 2015). Clusters were defined as the catchment area of a group of one to four community health workers working in the communities.

The clusters were allocated to one of four treatment groups: a control group, WaSH group, infant and young child feeding (IYCF) group, and combined WaSH and IYCF group. A constrained randomization technique was applied to achieve balance across all four arms for 14 key variables related to geography, demography, water access, and sanitation coverage. Recruitment into the trial was by prospective pregnancy surveillance and referral to the study of all eligible women found to be pregnant, and written informed consent was obtained to join the study. Community health workers delivered behavior change interventions; study staff measured outcomes.

In the SHINE study, distance to water and quantity of water used by each household were assessed through interviews and observation. However, water treatment at the point of use was provided to households in the experimental arms and tested in samples of drinking water collected at the household during home visits. An extensive list of covariates, as described in the SHINE statistical analysis plan (Ntozini et al., 2018), was collected from study participants by observation and interviews that includes maternal and infant characteristics, household characteristics, and environmental characteristics.

Study Sample and Sampling Procedures

The study sample is all participants who meet all inclusion criteria and are selected to participate in a study (Asiamah et al., 2017). The study sample is selected from the accessible population by applying further restrictions and then using a defined statistical sampling technique. For this research, the study sample is all children who participated in the SHINE trial and had a valid LAZ measurement at 18 months of age such that a stunting outcome could be assessed. The study sample in this case comprised almost the entire available accessible population as no sampling

was employed. This is one advantage of secondary data analysis—the low cost and time saving advantages allows for the analysis of all available data with no additional cost or time constraints (Smith et al., 2011).

Secondary Data Source and Variables

The analysis of secondary data is an established research methodology that can produce high impact research with advantages of lower costs and shorter time (Smith et al., 2011). Quantitative data collected during the conduct of the SHINE study between 2012 and 2017 in Chirumanzu and Shurugwi districts were used for answering research questions for this study. Permission to use de-identified data from the SHINE Study in the present analyses was sought from the principal investigator (see Appendix A). Variables extracted from the SHINE study for the present analysis are grouped into three groups: those that measure the exposure or access to water; those that measure the outcomes, which is linear growth; and those that explain the setting and context and are used as explanatory factors for adjusting the regression models during analyses.

Exposure Variables

Access to water was defined using three different metrics: quantity, quality, and accessibility. These were operationally defined as reported quantity of water used; the primary water source was protected or not; and reported distance or walk time to the primary water source. Each of these measured were defined at household level as well as at community or cluster level. Water was also classified according to intended purpose—either for drinking or for other household uses. Listed in Table 2 are the water access exposure variables collected in the SHINE trial that were required for this analysis.

Table 2

Water Access Exposure Variables Collected in the SHINE Study at Baseline (Antenatal)

Variable	Name	Description
1	HH_ID	Household unique ID
2	MOM_ID	Mom unique ID
3	CLUSTER_ID	Randomized cluster identifier
4	TOT_VOL	Total water volume per household per day
5	VOL_PER_CAPITA	Water volume per capita per day
6	DIST_DRINK_WATER	Distance to primary drinking water source
7	DIST_NONDRINK_WATER	Distance to primary nondrinking water source
8	TIME_DRINK_WATER	One way walk time to primary drinking water source
9	TIME_NONDRINK_WATER	One way walk time to primary nondrinking water source
10	DRINK_WATER_TYPE	Type of primary drinking water source
11	NONDRINK_WATER_TYPE	Type of primary nondrinking water source

Outcome Variables

Outcomes for the analysis were child anthropometric growth and nutritional status indicator, which were assessed at 18 months infant age. A summary of the variables collected is presented in Table 3.

Table 3

Variables Collected in the SHINE study at 18 Months Infant Age (Outcomes)

Variable	Name	Description
1	CHILDID	Child unique ID
2	MOM_ID	Mom unique ID
3	M18_VDATE	Date of 18 months visit
4	AGE_M18	Age at 18 months visit
5	M18_MHIV_STATUS	Maternal HIV status at 18 months
6	M18_IHIV_STATUS	Infant HIV status at 18 months
7	WEIGHT_M18	Weight (kg) at 18 months visit
8	LENGTH_M18	Length (cm) at 18 months visit
9	HC_M18	Head circumference at 18 months visit
10	MUAC_M18	Mid-Upper Arm Circumference (mm) at 18 months visit
11	LAZ_M18	Length-for-age z-score at 18 months visit
12	WAZ_M18	Weight-for-age z-score at 18 months visit
13	WHZ_M18	Wight-for-height z-score at 18 months visit
14	HCZ_M18	Head circumference z-score at 18 months visit

15	MUACZ_M18	Mid-Upper Arm Circumference z-score at 18 months visit
16	STUNTING	Stunting LAZ<-2
17	SEVERE_STUNTED	Severe stunted LAZ<-3
18	UNDERWEIGHT	Underweight WAZ<-2
19	WASTING	Wasting WHZ<-2
20	M18_DC	Data collector who assessed the child

Explanatory Variables

In the SHINE study dataset, a rich set of covariates that includes the randomized interventions, delivery and uptake of interventions, child growth measurements from birth to 18 months, maternal and household characteristics, and environmental variables were collected. These altogether were sufficient to answer the research questions. In Table 4 are the explanatory variables collected at the baseline visit.

Table 4

Explanatory Variables Collected in the SHINE Study at Baseline (Antenatal)

Variable	Name	Description
1	HH_ID	Household unique ID
2	MOM_ID	Mom unique ID
3	CLUSTER_ID	Randomized cluster identifier
4	DISTRICT	Study district
5	MOM_AGE	Mothers' age at enrollment
6	MOM_HEIGHT	Mom height (cm)
7	MOM_MUAC	Maternal Mid-Upper Arm Circumference (cm)
8	MOM_EDU	Education (highest class completed)
9	PARITY	Parity at baseline
10	MARRIED	Marital status
11	EMPLOYED	Mom employment status
12	MOM_MDDS	Mom meets dietary diversity
13	RELIGION	Religion
14	PREG_HIV_STATUS	Maternal HIV status during pregnancy
15	SES_SCORE	Baseline wealth index
16	CSI	Coping strategy index
17	ANYLATRINE	Any latrine at household?
18	IMPROVE_LATRINE	Improved latrine at household?
19	OPEN_DEF	Mom open defecation at house?
20	PROP_OPENDEF	Proportion open defecation at house?
21	FECESOBSERVED	Feces observed in the yard
22	IMPROVEDFLOOR	Improved floor at house?

23	TIMETOWATER	Time to the water point in minutes
24	TREATWATER	Treat water in any way
25	HHSIZE	Household size
26	FEMALE	Child sex
27	BIRTHWEIGHT	Infant birth weight (kg)
28	GEST_WEEKS	Gestation age in weeks
29	LOW_BW	Birth weight <2500g
30	PREMATURE	Premature (gestation age < 37weeks)

Data Analysis Plan

Data Exploration and Cleaning

Prior to analyzing the data, I first carried out data exploration and cleaning to check that the secondary data source met the requirements for answering the research questions. A major limitation of secondary data is that researchers may remove some key variables that they may deem to be sensitive before making the data available for public access to maintain confidentiality of their respondents (Cheng & Phillips, 2014). I checked the data for completeness, consistency, and the level of missing data in the responses. Data-cleaning is an important step to catch any errors in the data before any quantitative data analysis can be undertaken (Frankfort-Nachmias et al., 2015). I also checked the level of measurement scales used on key variables to ascertain if data were collected at the level of detail required for the analysis (Frankfort-Nachmias et al., 2015).

Descriptive Statistics

Descriptive statistics are an essential first step when conducting quantitative data analysis as they enable a research to understand their data better (Frankfort-Nachmias et al., 2015). In the study, descriptive statistics were generated to describe the study population; these included measures of central tendency and dispersion of key variables – mean and standard deviation (SD) for normally distributed continuous

variables, median and interquartile range (IQR) for continuous non-normal variables; proportions and frequency distributions for factor variables. Graphical summaries were also generated to explore univariable and bivariate relationships between variables. In addition, I generated maps to visualize the spatial distribution of some key exposure and outcome variables. Frankfort-Nachmias et al. (2015) asserted that presenting data graphically helps researchers visualize the data better to see any patterns and relationships. I analyzed the data using IBM PC SPSS version 25 to generate tables and graphs; Stata version 14.1 to perform the regression analyses; and Quantum Geographic Information Systems (QGIS version 2.8.3) to construct geospatial maps.

Inferential Statistics

Testing the Research Questions and Hypotheses. Regression analysis is one of the main quantitative methods used to test the nature of relationships between a dependent variable and one or more independent variables (Dudovskiy, 2018; Montgomery et al., 2012). Since the SHINE study was designed as a cluster randomized trial, data collected on child linear growth outcomes was also clustered around community health worker catchment areas. I therefore used regression methods that take into account this design to test the hypotheses. Generalized estimation equations (GEE) models are used to estimate population average changes in outcomes variables for longitudinal and clustered data (Wang, 2014). Wang described the conditions under which GEE models are used and the assumptions associated with their application. The general form of a GEE regression models is given in Equation 1.

$$g(\mu_{ij}) = \sum_i X'_{ij}\beta \quad \text{..... (1)}$$

Where: *g(.)* is a known link function for the response variable,

μ_i is the mean vector of observations Y_i in the *i*-th cluster,

X is an (n × p) vector of predictor variables x_{ij} where i = 1,...,n observations and j = 1,...,p predictors; and β is a (p × 1) vector of regression coefficients obtained.

The regression coefficients β are estimated by solving Equation 2:

$$\cup(\beta) = \sum_{i=1}^{K} D'_i V_i^{-1}(Y_i - \mu_i) = 0 \quad \text{..... (2)}$$

Where $D_i = \partial\mu_i/\partial\beta$' and V_i is the variance-covariance matrix for Y_i

Therefore, I used a GEE framework with appropriate link function and an exchangeable correlation structure to assess each hypothesis. I computed robust standard error estimates and their corresponding 95% confidence intervals to account for the clustered study design. I used the Wald Test statistic to assess model fit (Liu, 2015).

Model for Continuous Normally Distributed Response Variables. When the dependent variable was continuous (LAZ), I tested the associations with water access using the GEE models with a Gaussian distribution family, an identity link function *g(.)*, and an exchangeable correlation structure to estimate the mean difference in LAZ at various levels of the water access predictor variables as the exposure (see Equation 3).

$$g(\mu_{ij}) = \mu_{ij} = \sum_i X'_{ij}\beta \quad \text{..... (3)}$$

μ_i is the mean vector of normally distributed observations Y_i in the *i*-th cluster,

X is an (n × p) vector of predictor variables x_{ij} where i = 1,...,n observations

and j = 1,...,p predictors; and β is a (p × 1) vector of regression coefficients obtained that gives the predicted mean differences.

Model for Dichotomous Distributed Response Variables (Stunting). When the dependent variable was dichotomous such as childhood stunting, the variable was coded '1' when a child was stunted and '0' when not stunted. I tested the associations with water access using GEE models with a Binomial family and a log link function, also known as a log binomial regression model, to estimate relative risk (RR) of stunting at each level of the exposure. Log binomial models facilitate estimation of the RR as a measure of association (Huang, 2019). These models follow the general format shown in Equation 4.

$$g(\pi_{ij}) = \log(\pi_{ij}) = \sum_i X'_{ij}\beta \quad \text{..... (4)}$$

Where: π_i is the probability of stunting for the *i*-th observation

X is an (n × p) vector of predictor variables x_{ij} where i = 1,...,n observations, j = 1,...,p predictors; and β is a (p × 1) vector of regression coefficients obtained such that $e^{\beta j}$ gives the RR of stunting from the *j*-th predictor.

Model for Response Variables as a Proportion (Prevalence). When the dependent variable was a proportion such as percentage of stunted children in a cluster, which is bounded between 0 and 1, I tested the associations with water access using fractional regression models with a logit link function (Papke & Wooldridge, 1996; Woodrige, 2010). Fractional regression models enable estimation of percentage change in the outcome variable for each percentage or unit change in the predictor variable by estimating marginal effects as described in "Fractional outcome regression", StataCorp (2019). Fractional regression models follow the general form shown in Equation 5.

$$E[y \vee x] = e^{\sum_i X'_{ij}\beta} \Big/ 1 - e^{\sum_i X'_{ij}\beta} \text{...} \quad (5)$$

Where: y_i is the fractional outcome variable bounded between $0 \leq y \leq 1$; $logit(y) = y/(1-y) = X'\beta$; X is an $(n \times p)$ vector of predictor variables x_{ij} where $i = 1,...,n$ observations, $j = 1,...,p$ predictors; and β is a $(p \times 1)$ vector of regression coefficients. The percentage change in y relative to predictor x is then estimated by calculating the marginal effects, that is differentiating the likelihood function.

Testing the Study Hypotheses

I used different regression models as described above to test each of the hypotheses and answer the three research questions that follow:

- RQ 1: Is there an association between household water quantity and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi?
  - $H_0 1$: There is no association between household water quantity and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.
  - $H_a 1$: There is an association between household water quantity and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.
- RQ 2: Is there an association between household distance to water source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi?
  - $H_0 2$: There is no association between household distance/time to water source and child linear growth at the age of 18 months in rural

Chirumanzi and Shurugwi.

- *H_a2:* There is no association between household distance/time to water source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.

- RQ 3: Is there an association between household water quality at source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi?
  - *$H_0 3$:* There is no association between household water quality at source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.
  - *$H_a 3$:* There is no association between household water quality at source and child linear growth at the age of 18 months in rural Chirumanzi and Shurugwi.

For each of the three research questions, I tested associations between LAZ and water access using GEE models for continuous normally distributed outcomes and estimated mean differences. For associations between childhood stunting and water access, I used GEE models for binary outcomes and estimated RR. The independent variables were either factor variables such as “water quality” or continuous variables such as “distance to a water source”.

To test each hypothesis at community-level, I first generated community-level variables from the dataset using the following methodology: I computed summary measures of each exposure and outcome variable at the cluster-level and created a new dataset for the analysis. These were means, medians, or proportions depending on the distribution and form of the input variable. For normally distributed continuous

variables, I computed the mean value for each cluster; for continuous variables that had skewed distributions, I computed the median value for each cluster; and for binary variables I computed the mean value to estimate the cluster-level prevalence. I then estimated the community-level water access effects on cluster-level mean child linear growth and on the proportion of stunted children per cluster using linear regression analysis. I then employed fractional regression models with a logit link function and estimated percentage change in the outcome variable for each percentage or unit change in water coverage by estimating the marginal effects.

Adjusting for Confounding

Many study designs in epidemiology are susceptible to confounding due to the absence of randomization of the exposure of interest that makes them prone to many kinds of biases including selection bias, response bias, and recall bias among other biases (Creswell, 2009; Delgado-Rodriguez & Llorca, 2004; Hulley et al., 2007). In all regression analyses, I estimated univariable and multiple variable regression models and computed unadjusted and adjusted estimates of each measure of association and their corresponding 95% confidence intervals (95%CI) and *p-values*. To control for confounding, I selected explanatory variables from a list of available candidate variables that include maternal and infant demographic characteristics, household characteristics, and geographical or other structural characteristics could be associated with child growth and access to water.

Interpretation of Significance Levels

Two-sided *p*-values where used to declare statistical significance. The null hypotheses were not rejected when the associations were not statistically significant. *P*-values are derived on a continuous scale as a measure of the strength of evidence

against the null hypothesis H_0 (Ramsey & Schafer, 2012). I used the scale proposed by Ramsey and Schafer (2012) to rank the strength of evidence against the null hypothesis as either being very strong $p \leq 0.01$, strong $0.01 < p \leq 0.05$, and weak $0.05 < p \leq 0.10$ or no evidence of an association $p > 0.1$.

Treatment of Missing Values

Missing data can increase the risk of bias in observational research studies if they are not carefully assessed (Murray, 1998; Little & Rubin, 2014). Missing data mechanisms may not be random or independent of the outcome of interest (Murray, 1998). For example, in a survey where older respondents did not answer a question about exposure to a risk factor of a disease that has a higher risk in older people, then if analyses is performed on available responses from the younger respondents then this systematically missing response could underestimate the risk hence biasing the inference. Missing values were assessed and reported, if rates were lower (≤ 5%) then complete case analysis was conducted, however if missing data rates were higher (> 5%) then multiple imputation methods were considered in sensitivity analysis (Little & Rubin, 2014; Murray, 1998; Sterne et al., 2009). Multiple imputation, originally proposed by Rubin (1987) and further explained by Little and Rubin (2014) is a mechanism for filling in the missing data with plausible values but maintaining the variability in the responses and has become a preferred method of imputing missing data.

Effect Modification

The magnitude of the effect of an exposure on an outcome variable may vary depending on the presence of a third variable. The nature of this effect is called mediation, interaction, or effect modification (Corraini et al., 2019). I investigated the

presence of effect modification, interaction, and mediation for some key variables that could modify the effect of water access on childhood stunting interest in this study. From the published results of the SHINE study, infant exposure to maternal HIV was the main effect modifier that I included in the analyses. I tested the models for interaction between water access measures and maternal HIV-status. If the results showed evidence of an interruption ($p < 0.10$), I then presented results of each regression model stratified by maternal HIV-status.

Threats to Validity

Internal Validity

The internal validity of a study deals with the degree to which the results generated in the research are attributable to the independent variable and not any other possible explanation (Frankfort-Nachmias et al., 2015). Threats to internal validity of studies originate from multiple sources; measurements instruments used, sample selection procedures, and data collection methods. Threats to internal validity of research studies that perform secondary analysis of data also originate from the data sources used by inheriting the weaknesses of the original studies as well as introducing new ones. The current study will be based on secondary analysis of data collected in the SHINE Trial, a cluster randomized trial to test the efficacy of experimental interventions on child growth. Evidence from the implementation of the trial show high internal validity following extensive feasibility studies, pilot testing of the instruments, and regular standardization of data collectors (SHINE Trial Team, 2015). However, the threat to internal validity cannot be completely eliminated, but it can be minimized; for example, some data were collected from caregiver recall, which is prone to recall bias, then medical records were reviewed where possible to

triangulate the information. Distance to a water source, for example was collected as both time to water as well as distance to water source, and this allowed for validation of the metric. The current study is an observational study based on data from a clinical trial. Confounding is a serious threat to the internal validity of observational studies (Delgado-Rodriguez & Llorca, 2004). To increase the internal validity of this current observational study, all regression analysis will be adjusted for confounding using all available covariates that could be related to childhood stunting. I will also check for the presence of effect modification and interaction between variables; if found I will present results stratified according to the modifying variables.

External Validity

External validity is concerned about the degree to which results of a research study are applicable to the larger population. Threats to external validity may come from both the sample size and sampling methods used (Cheng & Philips, 2014). When a study lake power because of the sample size being too small, the results obtained may not be valid and cannot be inferred to be true to for the population. Likewise, if the sampling method selects a non-representative sample from the population, then results from such a study cannot be transferred to the target population (Delgado-Rodriguez & Llorca, 2004). To minimize selection bias and have sufficient power in this study, the entire SHINE study sample meeting inclusion criteria will be used for conducting this research. For any excluded participants, I will present a comparison of key variables between those included compared to those included in the analysis.

Missing data can increase risk of bias that can threaten the external validity of research studies if they are not carefully assessed (Murray, 1998). Missing data

mechanisms might not be random or independent of outcome of interest (Murray, 1998); for example, in a survey where older respondents systematically skip a question about exposure to a risk factor to a disease that has a higher risk in older people, then if such data were analyzed the results are likely to underestimate the risk in the target population. To minimize this threat in this study, missing values will be assessed and may be imputed if above certain threshold values as recommended by different researchers (Murray, 1998; Sterne at al., 2009).

Ethical Procedures

Protecting human participants is a mandatory step in any human participant research. To protect the research participants, their privacy, confidentiality, and freedom from potential risks must be always guaranteed. During the conduct of the SHINE trial, two Institutional Review Boards (IRBs) approved the original study (SHINE Trial team, 2015). For the current study, approval was obtained from the Walden University IRB to perform analysis of secondary data and answer the new research questions posed (IRB approval number 09-23-19-0582321). De-identified data was obtained from the Principal Investigator in the SHINE Study and used to perform these analyses (see Appendix A). Secondary analysis of data is normally performed on de-identified datasets to protect the participants' identities (Cheng & Philips, 2014). I also obtained Human Subject Research training in Good Clinical Practice and ICH (GCP) from the Collaborative Institutional Training Initiative (CITI Program ID: 37917231).

Summary

This quantitative analytic observational cohort study was designed to explore the association between childhood stunting and water access in rural resource limited

settings. The design and rationale of the study were discussed and the methodology explained. The data source was discussed, various threats to the validity of the study were explored and measures to address them explained. Procedures to ensure an ethical conduct of research on human participants were outlined. Finally, a detailed data analysis plan taking all the precautions was presented. The use of generalized linear models, including fractional regression and log-binomial regression analysis to assess the associations between dependent and independent variables, were discussed. In section 3 I present the results of my data analysis and findings of the study.

Section 3: Presentation of the Results and Findings

Introduction

The purpose of this prospective, observational, quantitative cohort study was to explore the association between childhood stunting and access to water in two districts in Zimbabwe. Access to water was operationally defined as three metrics: reported distance to water source; reported quantity of water used; and assessed quality of water used. The research three research questions assessed whether there was an association between household water quantity, household distance/time to water source, and household water quality at source and child linear growth at the age of 18 months in rural Chirumanzu and Shurugwi. Each of these exposure variables was defined at both household level and community or cluster levels. Water access was assessed for both drinking and non-drinking water sources.

Next, I briefly describe how the data were analyzed. I then present the study findings starting with summary descriptive statistics of the study participants and their settings and summaries of each exposure and outcome variable. This is followed by the main findings from each research question. The results include tables, graphs, and maps.

Data Analysis

I analyzed the data using IBM PC SPSS version 25 to generate tables and graphs, Stata version 14.1 to perform the regression analyses, and Quantum Geographic Information Systems (QGIS version 2.8.3) to construct geospatial maps for visualization. I conducted descriptive statistical analysis to describe maternal, infant, and household demographic characteristics and study settings. I then generated descriptive statistics of child growth (the study outcomes) as well as water access (the

study exposure variables) to describe the study population. I generated maps to illustrate the spatial distribution of child anthropometric measures as well as various water access metrics across the study area. To test the hypotheses at the child-level, I used GEE models to estimate population averaged effects of each exposure on the outcome and obtained robust standard error estimates. For the continuous outcome (LAZ) that is normally distributed, I used a Gaussian distribution family and identity link function to facilitate estimation of the mean difference in LAZ. For the binary outcome (child stunting), I used a binomial distribution family and log-link function to facilitate estimation of RR of child stunting. When the distribution of water access variables was skewed, I used a $\log_{10}$ transformation to stabilize the variance and bring the distribution of the transformed variable to near normality. I estimated the community-level water access effects on the mean child growth measures per cluster using linear and fractional regression analysis.

In all the analyses, I estimated the effect of water access on child growth in three steps: (a) The first model was not adjusted for any covariates; (b) the second model was minimally adjusted for confounding using study design factors of the original study, which include survey assessor, survey timing, and study randomized arms; and c) the third and final model was adjusted for confounding by putting in all the other measured covariates that could confound the effects of access to water on child growth. All covariates included in the analyses had < 5% missing values. I used the Wald Chi-Squared Test for assessing model fit (Liu, 2015). I tested the models for interaction between water access measures and maternal HIV-status, if results were statistically significant ($p < 0.10$) I then presented results of analysis stratified by maternal HIV-status.

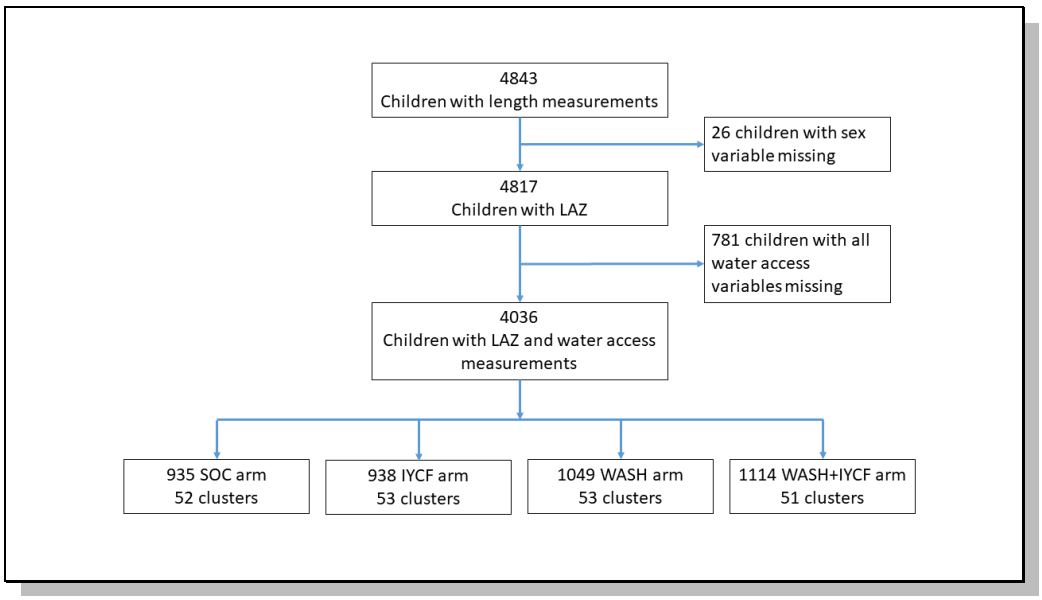
Study Results

Derivation of Study Sample

In Figure 4 is a summary of how the participants included in this analysis were selected from the secondary dataset analyzed. A total of *N* = 4,843 participants, recruited into the original study between 2012 and 2017, had child growth measurements at 18 months. Twenty-six children were excluded from analysis because their sex was missing from the dataset and a LAZ value; hence, stunting outcome could not be computed. An additional *n* = 781 participants did not have any water access measurements and were excluded from the analysis. This left *N* = 4,036 children in the final analysis. Study participants were drawn from a total of 209 randomized clusters with about 52 clusters in each study arm.

Figure 4

Derivation of Participants and Distribution by Randomized Study Arm and Clusters



Note. LAZ = length-for-age-z-score; SOC = standard of care; IYCF = infant and young child feeding; WASH = water, sanitation, and hygiene.

Characteristics of Study Participants

The baseline characteristics of the participants are summarized in Table 5. Mothers were mean ± SD: 26 ± 7 years, medium height 160 ± 6 cm, had good nutritional status, normal hemoglobin levels and had attained at least secondary education. Most of the mothers were unemployed. The majority were married, 40% met the minimum dietary needs, and 16% were HIV positive. The dominant religion was Apostolic, 45%. Twenty-two percent of the children were born premature and 8% were low birthweight. Only 36% of the women had an improved latrine at the household. There were 3 - 4% more women in the WASH and WASH+IYCF arms of the study compared to those in the SOC and IYCF arms.

Table 5

Explanatory Variables Collected in the SHINE Study at Baseline (Antenatal)

Continuous variables	n	Mean	SD	Median	IQR	
					LL	UL
Maternal age (years)	3863	26.3	6.6	25.7	20.7	31.3
Maternal height (cm)	3996	160	5.9	160	156	164
Maternal mid upper arm circumference (cm)	4022	26.4	3.0	26	24.3	28
Maternal hemoglobin (g/L)	3565	120	14.2	121	112	129
Maternal education (years)	3870	9.5	1.8	10	9	11
Parity	3243	1.9	1.4	1	2	3
Household size	3895	4.9	2.2	5	3	6
Infant birth weight (kg)	3703	3.1	.5	2.8	3.1	3.4
Factor variables	n	%				
Married (%)	3670	95.6				
Unemployed (%)	3670	91.3				
Wealth quintile (%)						
1 (lowest)	769	19.1				
2 (lower)	796	19.8				
3 (middle)	821	20.4				
4 (higher)	815	20.3				
5 (highest)	818	20.4				
Religion (%)						
Apostolic	1800	44.6				
Other Christian	1742	43.2				
Muslim and other	494	12.2				
Mother meets minimum	1550	39.5				

dietary diversity (%)		
HIV positive mothers (%)	645	16.0
Female children (%)	2013	49.9
Birthweight <2500g (%)	323	8.0
Premature, GA<37wks (%)	888	22.0
Any latrine (%)	1615	40.3
Improved latrine (%)	1421	35.5
Improved floor (%)	2158	53.5
Study arm (%)		
SOC	935	23.2
IYCF	938	23.2
WASH	1049	26.0
WASH+IYCF	1114	27.6

Note. N=4,036; SD=standard deviation; IQR=Inter-quartile range; LL=lower limit; UL=upper limit; GA=gestation age; SOC=standard of care; IYCF=infant and young child feeding; WASH=water, sanitation and hygiene.

Child Growth and Nutritional Status at 18 months

A summary of anthropometric measures of the children in the study at 18 months of age are summarized in Table 6. Children in the study, at 18 months were on average 1.6 ± 1.1 standard deviations below their expected height according to the WHO growth standard, 0.7 ± 1.0 standard deviations below their expected weight; however, they were not wasted or malnourished. WHZ and MUACZ were 0.03 ± 1.07 and 0.01 ± 0.89 respectively. A third of the children were stunted, 9% of them were severely stunted and 10% were underweight but only 3% were wasted.

Table 6

Child Growth and Nutritional Outcomes at Age 18 Months Among Study Participants

Continuous variables	n	Mean	SD
Child weight (kg)	4018	9.9	1.2
Child length (cm)	4036	77.7	3.2
Child head circumference (cm)	4007	46.6	1.6
Child mid upper arm circumference (MUAC) (mm)	4012	148	11
Child length-for-age z-score (LAZ)	4036	-1.57	1.09
Child weight-for-age z-score (WAZ)	4018	-.72	1.03
Child weight-for-height z-score (WHZ)	4014	.03	1.07
Child head circumference z-score (HCZ)	4007	-.26	1.09
Child mid upper arm circumference z-score (MUACZ)	4012	.01	.89

Factor variables	n	%
Child stunting, LAZ<-2 (%)	1343	33.28
Child severe stunting, LAZ<-3 (%)	366	9.07
Child underweight, WAZ<-2 (%)	418	10.40
Child wasting, WHZ<-2 (%)	114	2.84

Note. N=4,036; SD=standard deviation.

Sources of Water for Household Water and Their uses

Each household identified the primary sources of water that they used for drinking as well as for other household purposes. There were 15 different types of water sources. The distribution of water source types used for drinking and for other uses is shown in Figure 5. The most common sources for drinking water were boreholes, protected and unprotected deep wells, as well as unprotected shallow wells. About 9% of households used either surface water in rivers or dams or the river bed as their primary source of water for drinking. The predominant source of water for other household uses was surface water from a river, dam, stream, or lake. Over half of the participants (55%) used the same water source for both their drinking needs as well as other household water uses. However, among those who used an alternative water source for non-drinking purposes, more than half of the participants (54%) used unprotected surface water. There were virtually no piped water systems in all the households; only 1% of households had water piped into their plots and not into houses. The sources of water were further classified into four categories: piped on-plot, improved off-plot, unimproved ground water, and surface water according to their degree of protection and distance from the household. In Table 7, I summarize the distribution of the water sources according to their types and uses.

Table 7

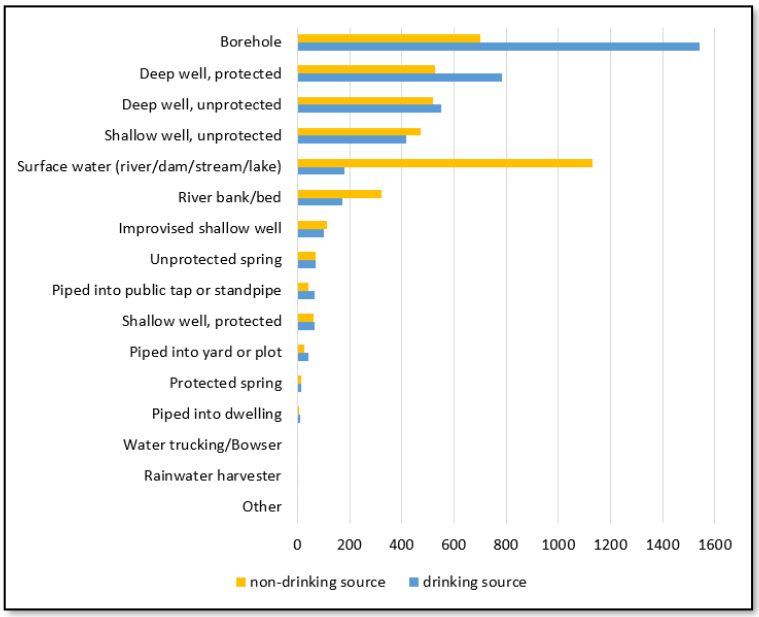
Distribution of Households According to Water Source Class and Their Uses

Type of primary water source	Drinking		Non-drinking	
	N	%	N	%
Piped on-plot	50	1.2	32	.8
Improved off-plot	2477	61.7	1341	33.5
Unimproved ground water	1140	28.4	1176	29.4
Surface water	351	8.7	1451	36.3

Note. N = 4,018.

Figure 5

Distribution of the Types of Water Sources by Use



Measures of Household Water Access

Household water access was measured in three ways as volume of water used, distance or walk time to a water source, and water source quality. The distribution of volumes of water collected were clustered in multiples of 20 liters as shown in Figure 6. The total volume of water fetched to the home by households in the last 24 hours was very variable and skewed to high with median 40 (IQR: 20-45) liters. Water volume used by each household member per-day was very limited with a median of 7

(IQR: 5-12) liters.

The average distance walked to a water source used for drinking was median 400 (IQR: 100-1,000) meters away from the homes, or a one-way walk time to the water source took median 10 (IQR: 5 to 20) minutes. The distribution of distance to water source used for drinking is shown in Figure 7. I also found a similar distribution for distance to non-drinking water sources.

Figure 6

Distribution of the Total Volume of Water Collected by Each Household Per-Day

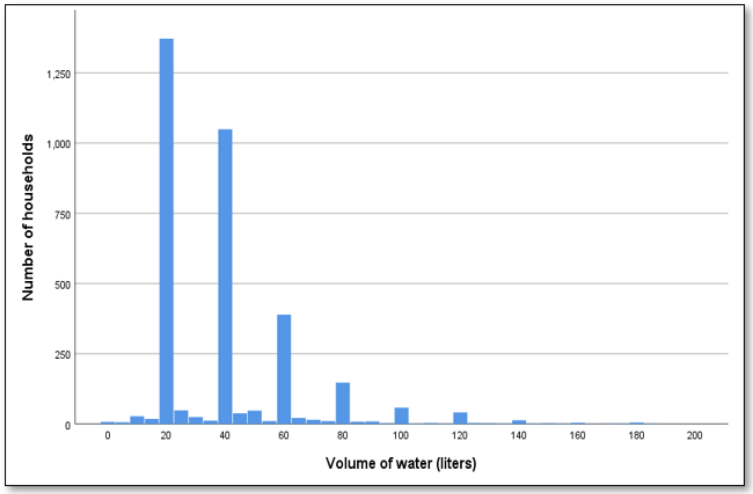
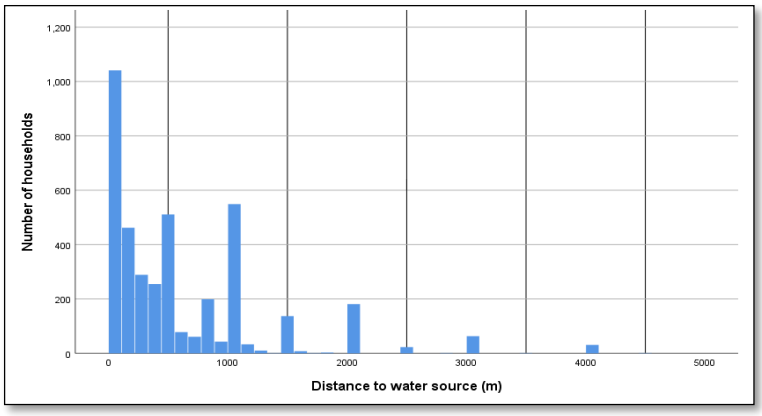


Figure 7

Distribution of the Distance of Households from the Primary Drinking Water Sources



The key measures of access to water and the quality of water used by households in the study are summarized in Table 8. About 26% of the households walked less than 100 meters to their primary drinking water source and 13% walked more than a kilometer. For 38% of households, one-way walk time was within 5 minutes; however, 7% of households walked more than 30 minutes to their primary water source. For most households, the primary water source for non-drinking purposes was located a similar distance to their primary sources for drinking water. Almost two thirds, 63% of households used protected water sources as their primary drinking water source, and 34% used protected water sources for others uses. Only 13% of households treated their drinking water in any way. Due to the skewed distribution of most of the water access variables, I used a $\log_{10}$-transformation to stabilize the variance and bring the distribution to approximate normality for the regression analyses. Summary measures of the transformed variables are also included in Table 8.

Table 8

Access to Water Among Households in the Study

Continuous variables	n	Mean	SD	Median	IQR	
					LL	UL
Total water volume fetched to the home in 24hrs (liters)	3441	42.7	46.7	40	20	45
Log_{10}(Total water volume)	3441	1.53	0.30	1.60	1.30	1.65
Water volume per person per day (liters)	3328	9.6	8.6	7.1	5	11.7
Log_{10}(Water volume per person per day)	3328	.88	.32	.85	.70	1.07
Distance to drinking water source (m)	4014	655	817	400	100	1000
Log_{10}(Distance to drinking water source)	4014	2.26	1.33	2.6	2	3
Distance to non-drinking water source (m)	4003	623	851	400	100	900
Log_{10}(Distance to non-drinking water source)	4003	2.18	1.37	2.60	2	2.95
Time to drinking water source	4012	14.8	16.4	10	5	20

(min)						
Log_{10}(Time to non-drinking water source)	4012	0.87	0.67	1	0.70	1.30
time to non-drinking water source (min)	4002	14.2	16.9	10	5	20
Log_{10}(time to non-drinking water source)	4002	0.83	0.68	1	0.70	1.30

Factor variables	n	%
Distance from household to water source used for drinking		
0-100m	1041	25.9
100-500m	1517	25.9
500-1000m	930	23.2
>1000m	526	13.1
One way walk time from household to water source used for drinking		
0-5 minutes	1,531	38.2
5-15 minutes	1,344	33.5
15-30 minutes	871	21.7
>30 minutes	266	6.6
Drinking water from a protected source (%)	2527	62.9
Water for other uses fetched from a protected source (%)	1373	34.3
Treat water in any way (%)	504	12.7

Note. N=4,036; SD=standard deviation. IQR= Inter-quartile range; LL=lower limit; UL=upper limit.

Predictors of Child Growth

Before testing the study hypotheses, I first investigated predictors of child growth that I could use as covariates in the multivariable regression models for testing each hypothesis. Variables that were statistically significant ($p < 0.1$) in univariable regression models predicting LAZ are shown in Table 9.

Table 9

Association of Maternal, Household and Infant Characteristics, and Child LAZ at 18 Months

Variable	N	Mean difference	95% CI		*P*
			LL	UL	
Mother HIV-status during pregnancy					
HIV-positive	654	reference	-	-	-
HIV-negative	3382	.317	.224	.410	<.001
Maternal height (cm)	3996	.051	.048	.056	<.001
Maternal MUAC (cm)	4022	.050	.040	.061	<.001

Mother employment status					
Not employed	3670	reference	-	-	-
Employed	348	-.073	-.206	.060	.284
Mother's religion					
Apostolic	1800	reference	-	-	-
Protestant	1742	.117	.050	.185	.001
Muslim and other	494	-.016	-.138	.107	.802
Improved household floor					
Basic mud floor	1812	reference	-	-	-
Improved floor	2158	.119	.046	.191	.001
SES quintile					
Lowest	771	reference	-	-	-
Lower middle	795	.121	.021	.221	.017
Middle	825	.150	.042	.257	.006
Upper middle	804	.293	.178	.407	<.001
highest	827	.290	.182	.398	<.001
Child low birth-weight					
>2500g	3381	reference	-	-	-
≤2500g	323	-.750	-.864	-.635	<.001
Child sex					
Male	2023	reference	-	-	-
female	2013	.305	.238	.372	<.001
Child born premature					
GA>37 weeks	3148	Reference	-	-	-
GA≤37 weeks	888	-.332	-.406	-.257	<.001

Note. CI=confidence interval; LL=lower limit; UL=upper limit; MUAC=mid-upper arm circumference; SES=socio-economic status; GA=gestation age.

Hypothesis 1: Effect of Water Quantity on Child Linear Growth

The quantity of water used in the home was first measured as a gross amount brought to a home in the last 24-hours. Secondly, per capita water volume was calculated by dividing the gross water quantity by the size of the household. I tested the association of the gross amount and per capita water volumes with child growth, LAZ and stunting at 18 months.

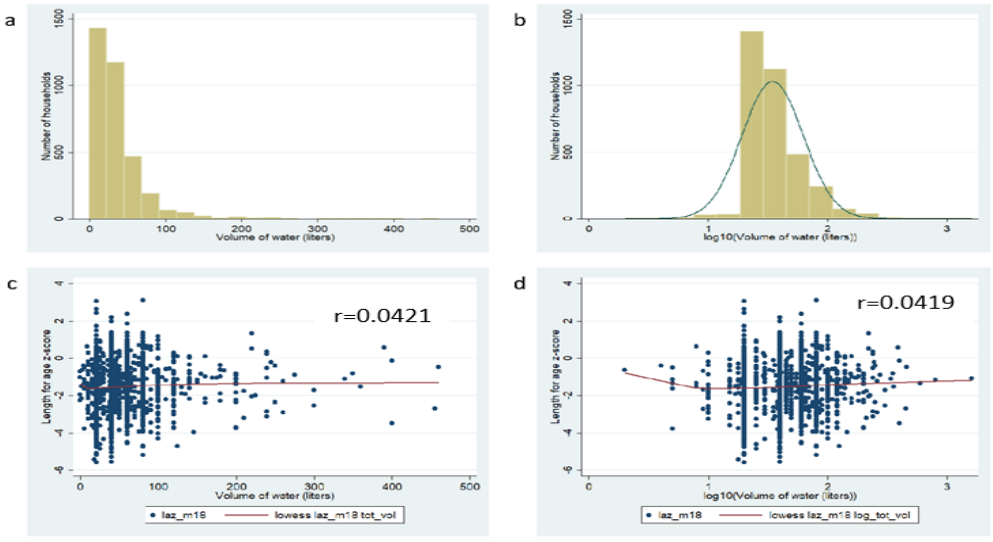
Exploring the Distribution of Total Water Volume on Child LAZ

In Figure 8, I used histograms and scatterplots to explore and visualize the association between gross water volume and child LAZ. I first plotted a histogram showing the

distribution of gross water volume shown in panel a, and a scatter plot of LAZ versus gross water volume shown in panel c. I then added a locally weighted scatterplot smoothing (LOESS) curve to visualize if there was a suggested trend between LAZ and gross water volume (Cleveland, 1979). The histogram shows that gross water volume is skewed to high with most frequent values of water volume lying between 0-100 liters and fewer households using up to 500 liters. The scatterplot shows a slightly positive gradient; however, the Pearson correlation coefficient (r = 0.042) shows no relationship, suggesting that LAZ may not be associated with volume of water used by a household. Due to the skewed distribution, I transformed gross water volume using a $\log_{10}$ transformation and plotted the semi-log histogram and scatter plot shown in Figure 8, panels b and d. The distribution of transformed water volume is closer to normality and the scatterplot values are distributed more evenly across the range of water volumes, however the correlation coefficient remained very weak (r=0.042) suggestion no association between LAZ and $\log_{10}$ water volume.

Figure 8

Gross Water Volume Per Day Used by Household and LAZ



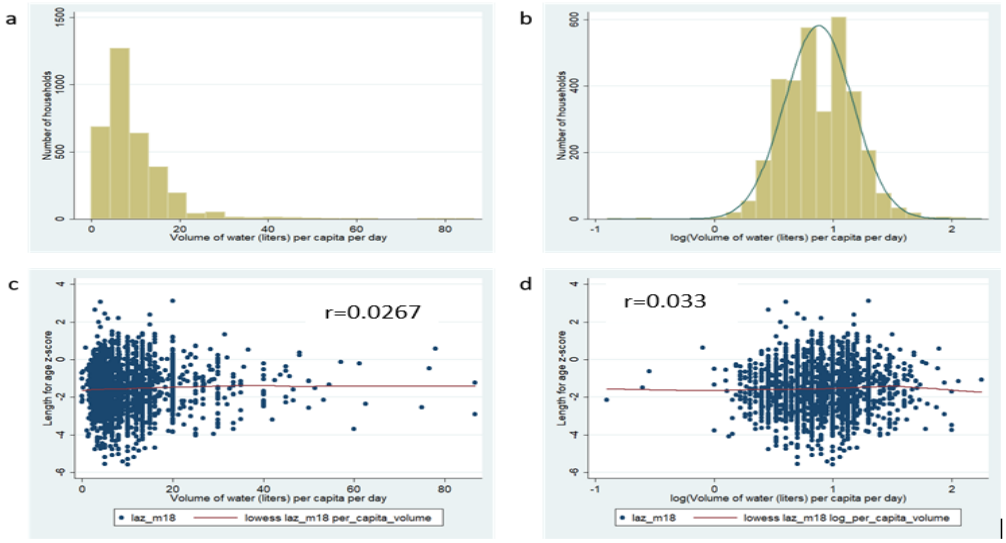
Note. Shown in the panels are a) gross water volume per day, b) $\log_{10}$ gross water volume with a normal distribution overlay, c) scatter plot of LAZ and gross water volume per day, and d) scatter plot of LAZ and $\log_{10}$ gross water volume per day used by household. The scatter plots have a locally weighted scatterplot smoothing (LOESS) curve overlay; **r** is the Pearson correlation coefficient.

Exploring the Distribution of Water Volume per Capita per Day on Child LAZ

Using a similar approach used for gross water volume, I plotted histograms and scatterplots of the original water volume per capita per day data (see Figure 9 panels a and c). The distribution of water volume per capita per day was skewed to high, hence I transformed the data using a $\log_{10}$ transform and the distribution shows a nearly normal distribution (see Figure 9, panel b). I then overlaid LOESS trend lines on the scatter plots (see panels c and d) that shows flat gradients and very weak Pearson correlation coefficients (r=0.027 and r=0.033) suggesting no association between LAZ and per capita water volume consumed by household members per day.

Figure 9

Per Capita Water Volume per Day Used by Household Members



Note. Shown in the panels are a) histogram of per capita water volume per day used by household members, b) a histogram of $\log_{10}$ per capita water volume with a normal overlay, c) a scatter plot of LAZ and per capita water volume, and d) a scatter plot of LAZ and $\log_{10}$ per capita water volume per day used by household. The scatter plots have a locally weighted scatterplot smoothing (LOESS) curve overlay; **r** is the Pearson correlation coefficient.

Association Between Water Volume and LAZ at 18 months

To test the association between water volume and child LAZ, I used GEE regression models to estimate the population-averaged changes in LAZ for each liter of water used. I conducted a three-stage analysis starting with and unadjusted model, then partially adjusting for trial related factors, and lastly the fully confounder adjusted analysis.

Unadjusted Analysis. I first ran four GEE models to estimate the unadjusted population averaged changes in LAZ for each measure of water volume, results are summarized in Supplementary Table B1. Total water volume was associated with gross water volume used by a household, mean LAZ increased by 0.001 (95%CI: 0.002, 0.017); $p = 0.01$, for each additional liter of water used. The results remained significant in the second GEE model using the $\log_{10}$ transformed gross water volume. The unadjusted model testing the association between per capita water volume and LAZ showed a weak association after the $\log_{10}$ transformation, mean LAZ changed by 0.106 (95% CI: -0.006; 0.218), $p = 0.06$, for each 1- $\log_{10}$ increase in liters per person per day.

Partially Adjusted Analysis. The second set of four minimally adjusted GEE models estimating population averaged changes in LAZ for each measure of water volume, results are summarized in Supplementary Table B2. In each of these models, I adjusted for the following covariates: Study arm – WASH and IYCF; data collector who took the measurement; and the calendar period of assessment. The above variables were factors that could have biased the measurement of each child's anthropometric measurement and are related to the design and conduct of the study that generated the data. Total water volume was strongly associated with LAZ; mean

LAZ increased by 0.001 (95%CI: 0.003, 0.017); $p = 0.007$, for each additional liter of water used. This association remained significant in the second model using the $\log_{10}$ transformed gross water volume. The minimally adjusted model testing the association between per capita water volume versus LAZ showed strong evidence of an association in the $\log_{10}$-transformed per capita water volume, the mean change in LAZ was 0.11 (95% CI: 0.0004; 0.220), $p = 0.05$.

Adjusted Analysis. I then conducted a third and final set of adjusted GEE models estimating the population averaged changes in LAZ for each measure of water volume; results are summarized in Table 10. In each of these models, I adjusted the models for baseline maternal, household, child, and environmental factors that were associated with the outcomes in univariable analyses (previously shown in table 6) as well as those related to the design of the study that generated the data. Total water volume was not associated with LAZ even after the $\log_{10}$ –transformation; the mean difference in LAZ was 0.0004 (95% CI: -0.0001; 0.001), $p = 0.14$. Water volume per capita per day showed strong evidence of an association with LAZ in the model using the $\log_{10}$ transformed water volume per capita per day. The mean change in LAZ was 0.100 (95% CI: 0.004; 0.196); $p = 0.04$, for each 1-$\log_{10}$ increase in liters per capita per day.

Table 10

Adjusted Association of Water Volume Used by Household and LAZ at 18 months

Water Access Variable	n	Adjusted mean difference	95% CI		*p*
			LL	UL	
Total water volume fetched to the home in 24hrs (liters)	3392	.0004	-.0001	0.001	.14
Log_{10}(Total water volume)	3392	.056	-.039	0.151	.25
Water volume per person per day	3281	.003	-.001	0.006	.16

(liters)					
Log_{10}(Water volume per person per day)	3281	.100	.004	0.196	**.04**

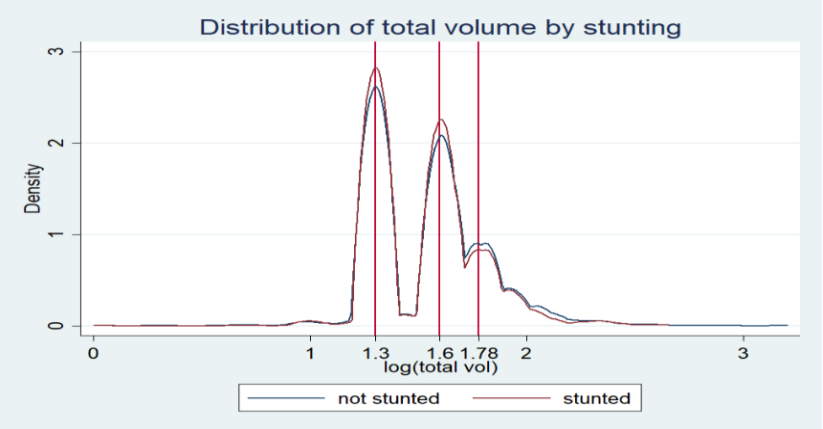
Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: HIV status, height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor; CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Exploring the Distribution of Water Volume and Child Stunting

The median volume of water used by households who had a stunted child was very similar to that used by households who had non-stunted children, median 40 (IQR: 20; 50) liters and 40 (IQR: 20; 40) liters respectively. In Figure 10, I present a Kernel Density Plot (Cleveland & Devlin, 1988), showing the frequency distribution of log_{10} transformed total water volumes used by households with stunted and non-stunted children. The shape of the distribution shows clumping around log_{10} = 1.30, 1.6, 1.78 corresponding to volumes of 20, 40, and 60 liters that is consistent with volumes carried in multiples of 20-liter plastic containers that are usually used by these households to bring water to their homes.

Figure 10

Distribution of Log_{10} Transformed Gross Water Volume Per Day Used by Households With a Stunted- and Non-Stunted Child



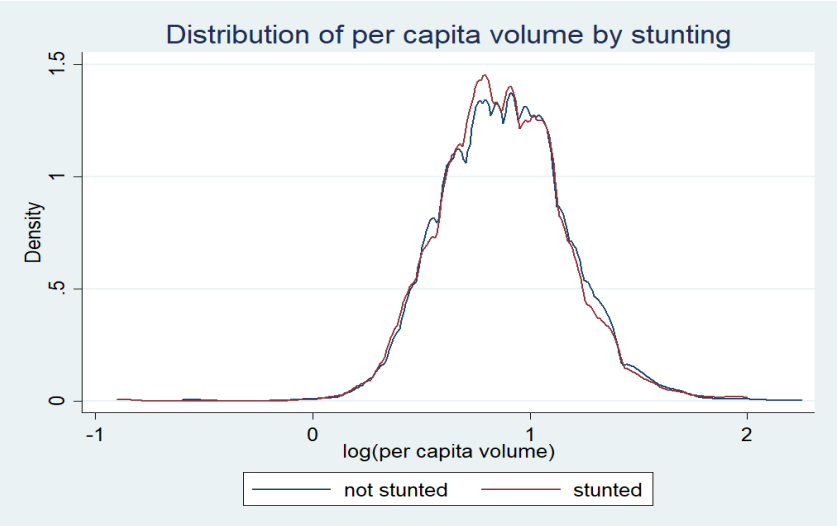
Note. Vertical lines mark volumes corresponding to 20, 40, and 60 liters.

Exploring the Distribution of Per Capita Water Volume and Child Stunting

I explored the characteristics of water volume per capita per day; the median volume of water used by households per capita per day was 7.5 (IQR: 5; 12) liters for households with non-stunted children while 6.7 (IQR: 5; 11.2) liters was used by households with a stunted child. Figure 11 illustrates a Kernel Density Plot showing the frequency distribution of $\log_{10}$ transformed per capita water volumes stratified according to child stunting. The plots show very similar distributions of water volume per capita among households with and without stunted children, which suggests that there may be no association between per capita water volume used by households and child stunting. I performed the formal testing of this association in the next section.

Figure 11

Distribution of $\log_{10}$ Transformed Water Volume Per Capita Per Day Used by Household Stratified by Child Stunting



Association Between Water Volume and Child Stunting at 18 months

To test the association between water volume and child LAZ, I used GEE regression models to estimate the population-averaged changes in LAZ for each liter

of water used. I conducted a three-stage analysis starting with and unadjusted model, then partially adjusting for trial related factors, and lastly the fully confounder adjusted analysis.

Unadjusted Analysis. I estimated the unadjusted RR using four GEE regression models, results are shown in Supplementary Table B3. In these models, only the untransformed total water volume was weakly associated with child stunting, unadjusted RR 0.999 (95% CI: 0.997; 1.000); $p = 0.05$, for each additional liter of water used. Water volume used per capita was not associated with child stunting in unadjusted analysis.

Partially Adjusted Analysis. I estimated population averaged RR of stunting using four GEE models, but this time the models were minimally adjusted for study design factors only, results are shown in Supplementary Table B4. The total volume of water was moderately associated with stunting at 18 months, adjusted RR 0.999 (95%CI: 0.997, 1.00); $p = 0.04$, for each additional liter of water used. This association however was not statistically significant in the $\log_{10}$-transformed regression model. Water volume per capita was not associated with child stunting in both partially adjusted regression models of untransformed and $\log_{10}$-transformed water volume measures.

Adjusted Analysis. In fully adjusted GEE regression models, results are shown in Table 11**;** both total water volume used and water volume usage per capita per day were not associated with child stunting at 18 months. The adjusted RR of stunting was 0.96 (95% CI: 0.82, 1.13); $p = 0.64$, for each $\log_{10}$-increase in total water volume used by a household. The adjusted RR for stunting was 0.93 (95% CI: 0.80, 1.07); $p = 0.30$, for each 1-$\log_{10}$ increase in per capita water volume usage per day.

Table 11

Adjusted Association of Water Volume Used by Household and Child Stunting at 18 Months Months

Water Access Variable	n	Adjusted RR	95% CI		*p*
			LL	UL	
Total water volume fetched to the home in 24hrs (liters)	3392	.999	.998	1.00	.25
Log_{10}(Total water volume)	3392	.96	.82	1.13	.64
Water volume per person per day (liters)	3281	.998	.99	1.00	.44
Log_{10}(Water volume per person per day)	3281	.93	.80	1.07	.30

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: HIV status, height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor. CI=confidence interval; LL=lower limit; UL=upper limit.

Interaction of Maternal HIV Status with Water Volume and Child LAZ

To assess if there were differences in the effect of water volume on child linear growth according to child HIV exposure, I tested the interaction ($p < 0.1$) between water volume used by the household and maternal antenatal HIV-status. Results of the tests for interaction of maternal HIV-status during pregnancy and water volume measures on child LAZ are summarized in Supplementary Table B5. There was a weak statistical interaction between maternal HIV status and total water volume used by a household; the difference-in-difference estimate (β) was -0.003 (95% CI: -0.007, 0.001); $p = 0.09$. This interaction suggested that the effect of water volume used by households on child linear growth may be different among children born to HIV-infected mothers compared to children not exposed to HIV. Due to the presence of a significant interaction, I analyzed the association of water volume and child LAZ stratified by maternal HIV-status.

Association Between Water Volume and Child LAZ among HIV-Unexposed Children

Unadjusted Analysis. I estimated the unadjusted effects of water volume on LAZ among children born to HIV-uninfected mothers, results are shown in Supplementary Table B6. Total volume of water fetched to the home in 24 hours was associated with increased child LAZ. This was significant in both the untransformed volume of water fetched as well as log_{10}-transformed measure. The unadjusted mean difference was 0.0007 (95% CI: 0.0003, 0.001); $p = 0.04$, per each additional liter used by a household and 0.12 (95% CI: -0.0028, 0.235); $p = 0.06$ per each 1-log_{10} increase in liters of water used by a household. Water volume used per capita per day was also weakly associated with increased child LAZ, unadjusted mean difference in log_{10} liters of water per person per day was 0.106 (95% CI: -0.011, 0.223), $p = 0.07$.

Adjusted Analysis. In Table 12, I summarize the adjusted effects of water volume on child LAZ among children born to HIV-uninfected women. Estimates from the GEE regression models adjusted for possible confounders were not statistically significant, all water volume measures were not associated with child LAZ among HIV-unexposed children.

Table 12

Adjusted Association between Water Volume Used by Household and LAZ at 18 Months Among HIV-unexposed Children.

Water Access Variable	N	Adjusted Mean difference	95% CI		*p*
			LL	UL	
Total water volume fetched to the home in 24hrs (liters)	2848	.0002	-.0003	0.0007	. 44
Log_{10}(Total water volume)	2848	.025	-.085	0.136	.65
Water volume per person per day	2751	.001	-.003	0.005	.54

(liters)					
Log_{10}(Water volume per person per day)	2751	.069	-.037	0.176	.20

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor; CI=confidence interval; LL=lower limit; UL=upper limit.

Association Between Water Volume and Child LAZ among HIV-Exposed Children

Unadjusted Analysis. I estimated the unadjusted and adjusted effects of water volume on LAZ among children born to HIV-infected mothers. In unadjusted analysis, results shown in Supplementary Table B7, total water volume fetched to the home in 24hrs was associated with increased child LAZ; unadjusted mean difference 0.004 (95% CI: 0.0004, 0.008); $p = 0.03$ for each additional liter used by a household. Water volume per person per day was also weakly associated with increased child LAZ; unadjusted mean difference was 0.013 (95% CI: -0.002, 0.029); $p = 0.09$ for each additional liter used by a person per day. The log_{10} transformed water volume measures were not significantly associated with child linear growth.

Adjusted Analysis. In multivariable GEE models, Table 13 shows the adjusted effects of water volume on LAZ among children born to HIV-infected mothers. Total water volume fetched to the home in 24 hours was weakly associated with increased child LAZ with adjusted mean difference of 0.003 (95% CI: 0.0001, 0.007); $p = 0.06$ per each additional liter of water used by a household. Water volume per person per day was strongly associated with increased child LAZ in adjusted analyses; adjusted mean difference 0.014 (95% CI: 0.002, 0.025); $p = 0.02$ per each additional liter used by a person per day. Log_{10}-transformed volume measures were not statistically significant in adjusted GEE regression models.

Table 13

Adjusted Association of Water Volume Used by Household and LAZ at 18 Among Months Among HIV-Exposed Children

Water access variable	n	Adjusted mean difference	95% CI		*p*
			LL	UL	
Total water volume fetched to the home in 24hrs (liters)	544	.003	-.0001	.007	**.06**
Log_{10}(Total water volume)	544	. 59	-.141	.450	.30
Water volume per person per day (liters)	530	.014	.002	.025	**.02**
Log_{10}(Water volume per person per day)	530	.175	-.103	.452	.22

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor. CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Hypothesis 2: Effect of Distance of Household from Water Source on Child Linear Growth

Available data included distance of each household from the primary water sources used for both their drinking needs as well as for other household uses. Therefore, I tested the hypotheses that distance to a primary drinking and non-drinking water source was associated with child LAZ and stunting at 18 months.

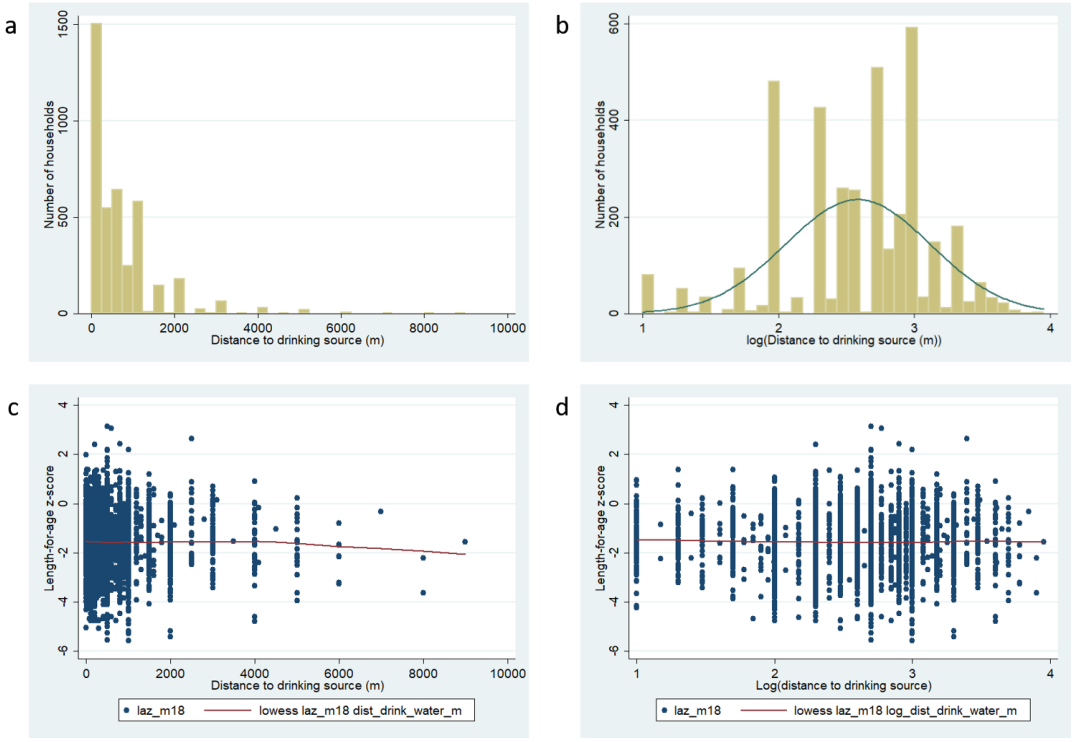
Exploring the Distribution of Distance to Primary Water Source and LAZ

In Figure 12 I show graphs summarizing the distribution of distance to drinking water source. The histogram in panel A shows that distance to drinking water source has a positive skewness. The histogram of the log_{10}-transformed distance to drinking water source (panel B) is closer to a normal distribution. The LOESS curve on the scatter plot in panel C shows that the distribution of distance to drinking water source and child LAZ has a slight negative correlation meaning that LAZ decreases

with increasing distance from a source of drinking water. However, the trend in the $\log_{10}$-transformed distance graph in panel D is almost flat suggesting that the association between distance to drinking water source and child linear growth may be very weak. I repeated the same exploratory analysis on distance to primary water source used for non-drinking purposes (graph not shown). I found that the distribution of distances was very similar to that of drinking water and the trend between child LAZ and distance to non-drinking water source was very similar to that seen with drinking water.

Figure 12

Distance to Drinking Water and LAZ and Distance to Primary Water Source Used for Drinking by Household Members



Note. Panels (a) shows a histogram of distance to drinking water source; (b) shows a histogram of $\log_{10}$ distance to drinking water source; (c) shows a scatterplot of LAZ and distance to water source in meters; (d) shows scatterplot of LAZ and $\log_{10}$ distance to water source.

Association Between Distance to Primary Water Source and Child LAZ

To test the association between distance to water source and LAZ, I used GEE regression models to estimate the population-averaged changes in LAZ for each measure of distance to water source. I conducted a three-stage analysis starting with and unadjusted model, then partially adjusting for trial related factors, and lastly the fully confounder adjusted analysis.

Unadjusted Analysis. Results of unadjusted analysis are shown in Supplementary Table B8; child LAZ at 18 months was not associated with distance to a primary water source for drinking; the mean difference in LAZ was -0.010 (95%CI: -0.036, 0.016); $p = 0.47$ for each 1-$\log_{10}$ increase in distance in meters walked. Child LAZ at 18 months was however weakly associated with distance to a primary source of water for other non-drinking purposes, the mean difference in LAZ was -0.023 (95%CI: -0.048, 0.003); $p = 0.09$ for each 1-$\log_{10}$ increase in distance in meters walked.

Partially Adjusted Analysis. I estimated the population-averaged changes in LAZ for each measure of distance to water source using partially adjusted GEE models accounting for the design of the trial, results shown in Supplementary Table B9. In each of the four GEE models, I adjusted for the same trial factors: study arm – WASH and IYCF; the data collector who took the measurement; and calendar period of assessment. Distance to a primary water for either drinking or for other uses was not associated with child LAZ at 18 months in any of the regression models.

Adjusted Analysis. In fully adjusted GEE models estimating the adjusted population-averaged changes in LAZ for each measure of distance to water source; results are summarized in Table 14. In each of these regression models, the factors

that I adjusted for are noted in the footnote of each table. The only statistically significant result was distance to water for non-drinking purposes with adjusted mean difference in LAZ of 0.00003 (95% CI: -0.0000, 0.0001); $p = 0.08$. However, this weak association was not significant in the regression model using the $\log_{10}$-transformed distance to source suggesting it might be a spurious result.

Table 14

Adjusted Association of Distance to Primary Water Source and LAZ at 18 Months

Water access variable	n	Adjusted mean difference	95% CI		*p*
			LL	UL	
Distance to primary water source for drinking (m)	3954	.0000	-.0000	.0001	.57
Log_{10}(Distance to primary water source for drinking)	3954	.0068	-.0170	.0307	.58
Distance to primary water source for other uses (m)	3943	.00003	-.0000	.0001	**.08**
Log_{10}(Distance to primary water source for other uses)	3943	.0008	-.0226	.0242	.95

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: HIV status, height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor. CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Exploring the Distribution of Distance to Primary Water Source and Child Stunting

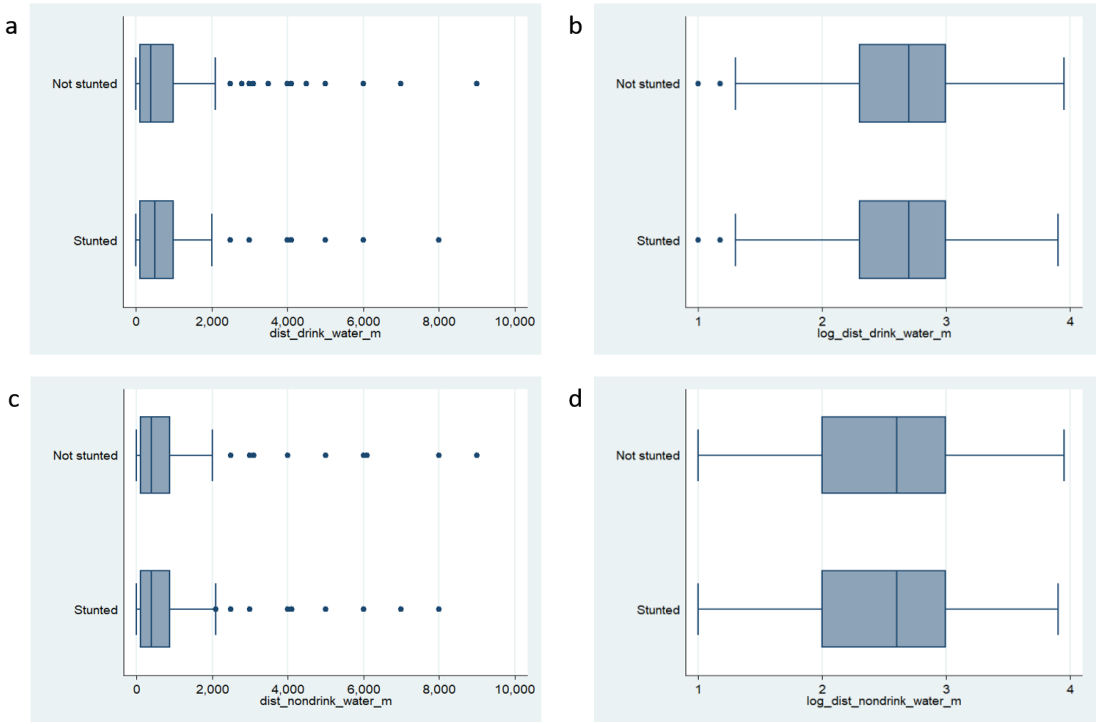
I examined summary descriptive statistics of distance from household to primary water sources according to presence of stunted child in the home. Households with a stunted child walked slightly further to fetch drinking water than households with non-stunted children, median 500 (IQR: 100; 1,000) and 400 (IQR: 100; 1,000) respectively. Distance to non-drinking water sources was similar between both households with and without a stunted child.

Figures 12 shows boxplots illustrating the distribution of distance to primary water source for drinking and for other uses respectively. Panels A and C shows the

distribution of distance in meters to a drinking and non-drinking water sources respectively, according to presence of stunted child in the home. The distributions are skewed to high, hence I also plotted the distribution of distance to water sources after the $\log_{10}$-transformation that are more symmetric (panels B and D). Even though most households were within one km of their primary water source, some households travelled up to 10 km to fetch their water. In the next section, I then tested these associations using regression models.

Figure 13

Boxplots of Distance to Primary Water Sources by Child Stunting



Note. Panel (a) distance to drinking water source, and (b) $\log_{10}$ transformed distance to drinking water source. Panel (c) distance to primary non-drinking water source, and (d) $\log_{10}$ transformed distance to primary non-drinking water source.

Association of Distance to Primary Water Source and Child Stunting at 18 months

I conducted the three-stage analysis starting with and unadjusted model, then partially adjusting for trial related factors, and finally running the fully confounder

adjusted analysis.

Unadjusted Analysis. I first estimated unadjusted RR of stunting for each measure of distance to water, results shown in Supplementary Table B10. Distance to water source was not associated with child linear growth for either drinking purposes or other household uses, the 95% confidence interval around each unadjusted RR estimate included 1.0 for all the GEE regression models.

Partially Adjusted Analysis. I then estimated partially adjusted GEE regression models adjusted only for study related factors only, results shown in Supplementary Table B11. Again, none of the regression estimates were significant as all the 95% confidence intervals around each estimated adjusted RR included 1.0 and *p*-values > 0.1.

Adjusted Analysis. In fully adjusted GEE models, the distance of household to their primary water sources for any uses, drinking or other uses, was not associated with child stunting at 18 months, results are shown in Table 15. The analysis shows that risk of childhood stunting does not depend on the distance walked to fetch water, even after taking into account all possible confounders that were measured.

Table 15

Adjusted Association of Distance to Primary Water Source and Child Stunting at 18 Months

Water access variable	n	Relative risk of stunting	95% CI		***p***
			LL	UL	
Distance to primary water source for drinking (m)	3954	1.00	1.00	1.00	.74
Log_{10}(Distance to primary water source for drinking)	3954	.996	0.96	1.03	.79
Distance to primary water source for other uses (m)	3943	.99999	.9999	1.00	.66
Log_{10}(Distance to primary water	3943	.99	.96	1.03	.71

source for other uses)

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: HIV status, height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor. CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Interaction Between Maternal HIV-Status with Distance to Primary Water Source and Child LAZ

Exposure to maternal HIV did not modify the association between LAZ and distance to primary water sources for any use. I did not find any statistically significant interactions between maternal HIV status and any measure of distance to primary water source used by households in the study; results are summarized in Table 16. Hence, I did not conduct stratified results in this analysis.

Table 16

Interaction Between Distance to Primary Water Source and Maternal HIV status on Child LAZ at 18 months

Water access variable	n	β	95% CI		*p*
			LL	UL	
Maternal HIV status × Distance to primary water source for drinking (m)	4014	-.00004	-.0002	.0001	.53
Maternal HIV status × Log_{10}(Distance to primary water source for drinking)	4014	-.037	-.116	.042	.36
Maternal HIV status × Distance to primary water source for other uses (m)	4003	-.00004	-.0002	.0001	.49
Maternal HIV status × Log_{10}(Distance to primary water source for other uses)	4003	-.035	-.106	.037	.34

Note. β = Difference-in-difference estimate; CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face. Interaction $P<0.10$ was considered as evidence of interaction.

Hypothesis 2a: Effect of Walk-time to Primary Water Source on Child Linear Growth

I first explored using graphics the bivariate relationships between water access

measures and each child growth outcomes. I then conducted regression analyses using both walk-time to drinking and walk-time to non-drinking water sources as predictors of child linear growth and stunting.

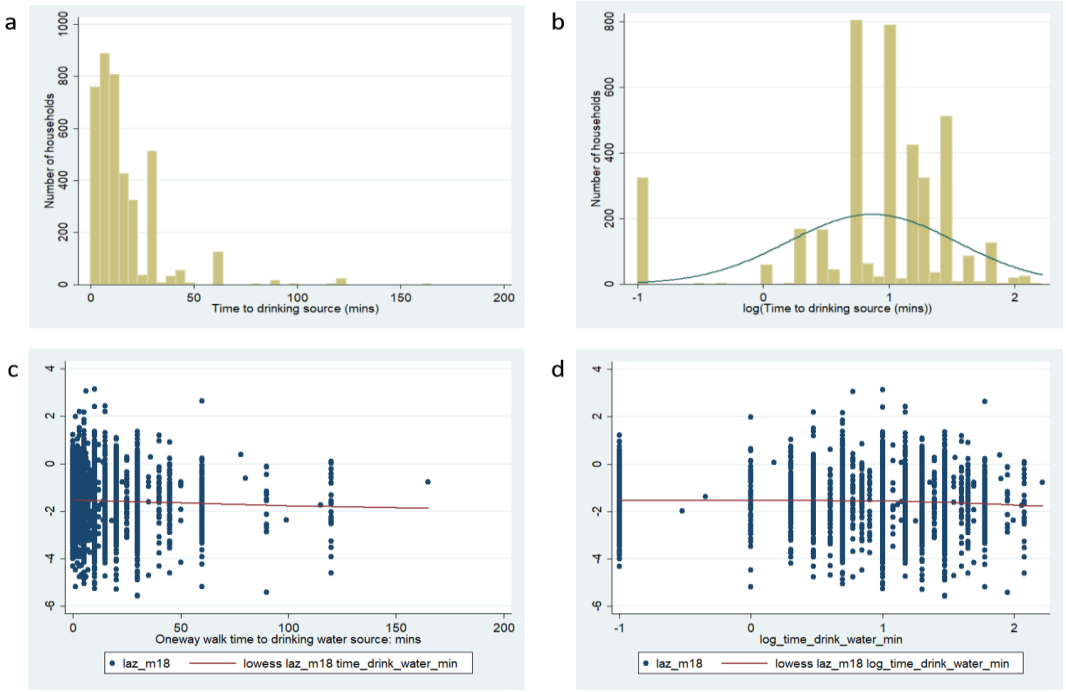
Exploring the Distribution of One-Way Walk-Time to Primary Water Source and LAZ

To explore the association between walk-time to water and LAZ, I first plotted scatterplots showing the distribution of one-way walk-time to drinking water source and child LAZ shown in Figure 14. The graphs shown shows the distribution of walk-time in minutes (panel a) that is skewed to high, and in panel b the distribution of $\log_{10}$ transformed walk-time that is nearly normally distributed. I then plotted a LOESS curve on the scatterplot to visualize the trend between LAZ and walk-time in minutes. The plot exhibits a slight negative gradient meaning that LAZ decreases with increasing walk-time to a drinking water source used by a household.

I repeated the same analysis this time exploring walk-time to primary water source used for non-drinking purposes by households, and plotted graphs similar to Figure 14 (not shown). There was very little variation in LAZ as walk-time to non-drinking water source increases that may mean that the association between child linear growth and walk-time taken fetching water for non-drinking purposes could be very weak. I then proceeded to test these associations using regression models.

Figure 14

Histograms and Scatterplots of LAZ and One-Way Walk-Time to Primary Water Source Used by Household Members



Note. Panel (a) shows histograms of walk-time to drinking water source in minutes; (b) histogram of $\log_{10}$ walk-time to drinking water source; (c) scatterplot of LAZ and walk-time to drinking water source in minutes; and (d) scatterplot of LAZ and $\log_{10}$ walk-time to drinking water source.

Association Between One-Way Walk-Time to Primary Water Source and Child LAZ

I conducted a three-stage analysis starting with and unadjusted model, then partially adjusting for trial related factors, and lastly the fully confounder adjusted analysis.

Unadjusted Analysis. I first conducted an analysis using GEE models to estimate the unadjusted population averaged changes in LAZ for each measure of walk-time to a water source, results shown in Supplementary Table B12. The only significant model was for LAZ and one-way walk-time to primary water source for drinking; unadjusted mean difference in LAZ of -0.002 (95% CI: -0.005, -0.000) and

$p = 0.05$. However the log transformed walk-time did not reach statistical significance; mean difference in LAZ of -0.041 (95% CI: -0.091, 0.009); $p = 0.11$. Time to non-drinking water was not associated with child linear growth.

Partially Adjusted Analysis. I conducted the second stage of my analyses estimating a set of GEE models partially adjusted for trial design factors. None of the associations were statistically significant, results shown in Supplementary Table B13. The model for one-way walk-time to primary water source for drinking, which was significant in unadjusted analyses had a mean difference in LAZ of -0.002 (95% CI: -0.004, 0.001); $p = 0.14$.

Adjusted Analysis. Lastly, I conducted the third stage of my analysis by estimating a set of GEE models fully adjusted for any confounding, and the estimated population averaged changes in LAZ for each measure of walk-time to a water source are summarized in Table 17. The only association reaching a very weak statistical significance level was one-way walk-time to primary water source for other uses; mean LAZ was 0.002 (95% CI: -0.0003, 0.003); $p = 0.10$. None of the other associations tested were statistically significant.

Table 17

Adjusted Association of Walk-Time to Primary Water Source and LAZ at 18 months

Water access variable	n	Adjusted mean difference	95% CI		*p*
			LL	UL	
One-way walk-time to primary water source for drinking (mins)	3953	-.001	-.003	.002	.66
Log_{10}(One-way walk-time to primary water source for drinking)	3953	.010	-.036	.056	.67
One-way walk-time to primary water source for other uses (mins)	3943	.002	-.0003	.003	.10
Log_{10}(One-way walk-time to primary water source for other uses)	3943	.026	-.020	.071	.27

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: HIV

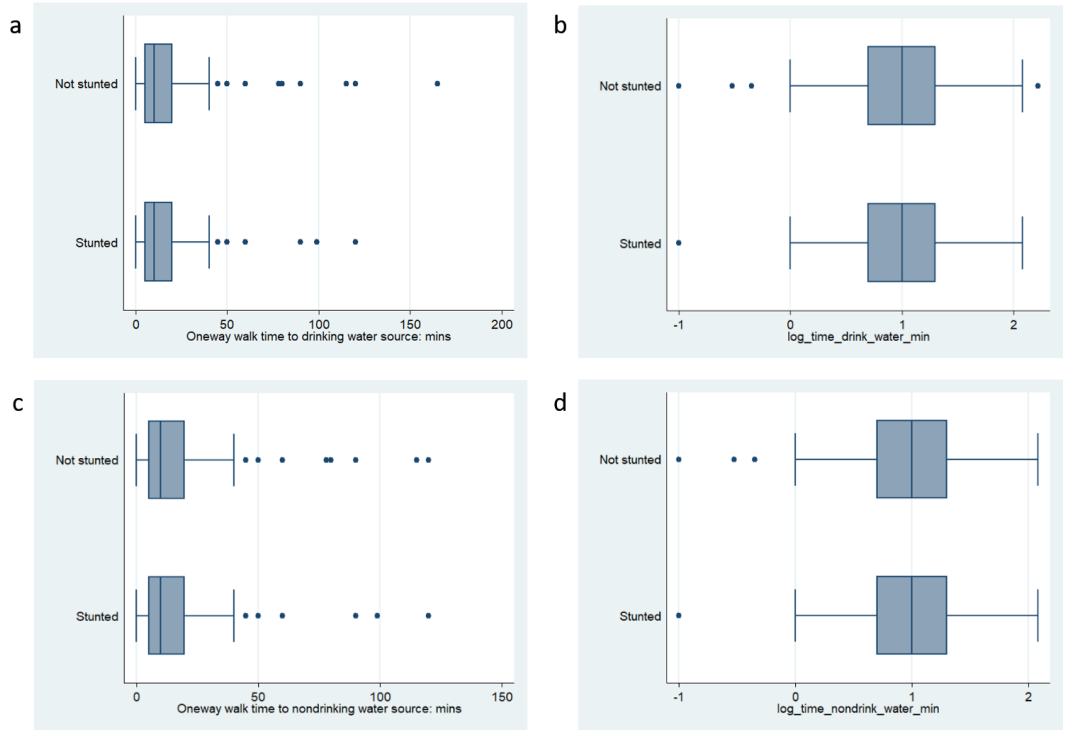
status, height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor; CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Exploring the Distribution of One-Way Walk-Time to Primary Water Source and Child Stunting

A summary of walk-time to primary water source is given in Figure 15. There were no observed differences in either one-way walk-time to primary water sources for drinking or non-drinking among households with stunted and stunted children. The median time to the water source was 10 minutes (IQR: 5, 20) among both stunted and non-stunted household to any water source.

Figure 15

Boxplots of One-Way Walk-Time to Primary Water Sources Used by Household For Drinking and Non-Drinking by Child Stunting



Note. Panels (a) shows the distribution of walk-time to drinking water; (b) shows distribution of $\log_{10}$ walk-time to drinking water source; (c) shows the distribution of walk-time to non-drinking water source; and (d) shows distribution of $\log_{10}$ walk-time to non-drinking water source.

Association of One-Way Walk-Time to Primary Water Source and Child Stunting At 18 Months

I conducted a similar three-stage analysis starting with and unadjusted model, then partially adjusting for trial related factors, and lastly the fully confounder adjusted analysis.

Unadjusted Analysis. In unadjusted GEE models estimating the RR of stunting for each measure of walk-time to a primary source of water, results shown in Supplementary Table B14; only one-way walk-time to primary water source for drinking was weakly associated with child stunting; unadjusted RR 1.00 (95% CI: 1.00, 1.01); $p = 0.07$. However, when walk-time was $\log_{10}$ transformed the model did not research statistical significance; unadjusted RR 1.04 (95% CI: 0.97, 1.11); $p = 0.29$. Walk-time to the primary water source for non-drinking was not associated with child stunting.

Partially Adjusted Analysis. In the second set of models that I estimated, partially adjusting for trial related factors, results shown in Supplementary Table B15; – none of the measures of walk-time to primary water source were associated with child stunting; walk-time to any primary water source for either drinking or non-drinking purposes was not associated with child stunting after adjusting for study design structural factors.

Adjusted Analysis. Similarly, in fully adjusted analysis, results shown in Table 18, none of the associations estimated by the GEE models were statistically significant. One-way walk-time to a primary water source for all purposes – drinking or non-drinking was not associated with child stunting after adjusting for the measured potential confounders.

Table 18

Adjusted Association of Walk-Time to Primary Water Source and Child Stunting At 18 Months

Water access variable	n	Adjusted RR	95% CI		*p*
			LL	UL	
One-way walk-time to primary water source for drinking (minutes)	3953	1.00	.998	1.00	.71
Log_{10}(One-way walk-time to primary water source for drinking)	3953	.98	.92	1.05	.62
One-way walk-time to primary water source for other uses (minutes)	3943	.999	.997	1.00	.56
Log_{10}(One-way walk-time to primary water source for other uses)	3943	.97	.90	1.04	.36

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: HIV status, height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor. CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Interaction Between Maternal HIV-Status with Time to Primary Water Source On Child LAZ

The effect of walk-time to water source on the risk of childhood stunting was not modified by infant exposure to maternal HIV. I did not find any statistically significant interactions between maternal HIV status and any measure of one-way walk-time to primary water source used by households in the study. Hence, I did not conduct stratified analyses for this hypothesis.

Hypothesis 3: Effect of Primary Water Source Quality on Child Growth

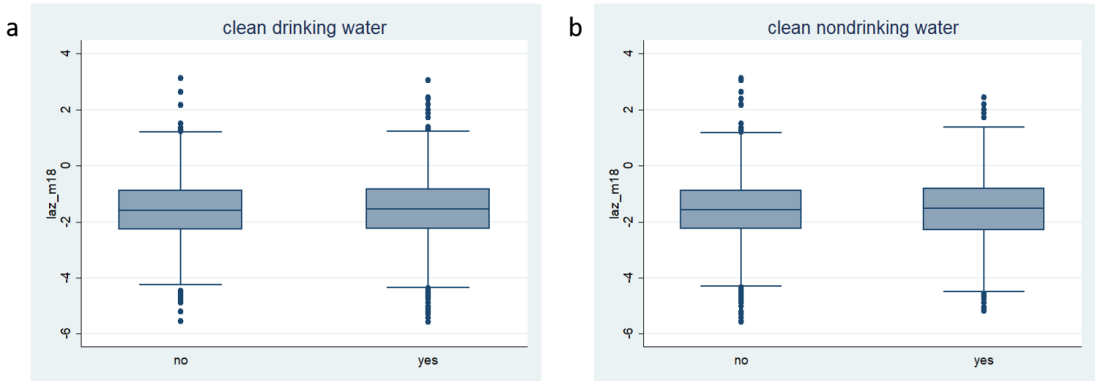
The water quality of each primary drinking and non-drinking water sources for the household was assessed. Water quality was classified as either safe if the source was protected or unsafe if the water source was unprotected. I tested the hypotheses that water quality was associated with LAZ and child stunting at 18 months, testing the associations for both drinking and non-drinking water sources separately.

Exploring Water Source Quality and Child LAZ at 18 Months

I summarized the descriptive statistics and the distribution of child LAZ by water source quality. Children from households with protected water sources used for any use had higher mean LAZ by between 0.04 and 0.05 standard deviations. The boxplots depicted in Figure 16 however shows that the distributions of LAZ were very similar between groups.

Figure 16

Boxplots of LAZ by Water Source Quality



Note. Panel (a) shows the distribution of LAZ by drinking water quality, and (b) shows distribution of LAZ by non-drinking water quality.

Association of Water Source Quality and LAZ at 18 Months

Following the three-stage analysis approach where I start with unadjusted models, then partially adjusting for trial related factors, and finally fully confounder adjusted analysis, I tested the association between LAZ and water source quality.

Unadjusted Analysis. Results of unadjusted GEE models estimating the population averaged mean difference in LAZ for each type of water source are showed in Supplementary Table B16; none of the associations between LAZ and water source quality were statistically significant. The estimated unadjusted mean difference in LAZ for children from households with protected water sources was

0.04 (95% CI: -0.04, 0.12); $p = 0.32$, and the estimated effect size for non-drinking water sources was even less.

Partially Adjusted Analysis. After partially adjusting the estimates for study design related factors, results shown in Supplementary Table B17; estimated effect sizes were very similar and associations between LAZ and water source quality were not significant. Water source quality was not associated with child linear growth after partially adjusting for trial design factors.

Adjusted Analysis. The quality of water for any household use was not associated with child linear growth after fully adjusting for confounding and trial design related factors (see Table 19). The estimated effect size of the adjusted mean difference in LAZ between children from households with protected and unprotected water sources were very similar and not statistically significant.

Table 19

Adjusted Association of LAZ at 18 Months with Quality of Primary Water Sources

Water access variable	n	Adjusted mean difference	95% CI		*p*
			LL	UL	
Primary source of water for drinking is protected	3958	-.03	-.09	.04	.44
Primary source of water for other uses is protected	3940	-.02	-.09	.05	.55

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: HIV status, height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor. CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Exploring Water Quality and Child Stunting at 18 months

The prevalence of stunting was marginally lower among children from households with access to protected water sources compared to children with access

to unprotected water sources. The proportion of stunted children from households with access to protected water was 34% (514/1,491) and 32% (821/2,527) in households with access to unprotected water sources.

Association of Water Quality and Child Stunting at 18 Months

Unadjusted Analysis. Starting with unadjusted GEE models, I tested the association between stunting and water source quality to estimate unadjusted RR of stunting, results shown in Supplementary Table B18. When compared to children from households with protected sources of drinking water, children from households with unprotected sources of drinking water had similar risk of stunting – unadjusted RR 0.95 (95% CI: 0.86, 1.05); $p = 0.29$ and 0.98 (95% CI: 0.89, 1.08); $p = 0.69$ respectively. None of the associations were however statistically significant.

Partially Adjusted Analysis. After partially adjusting the GEE models for study design factors, results shown in Supplementary Table B19; the estimates did not change much from the unadjusted effect sizes. Protected water sources were associated with a slightly reduced risk of stunting, however the effect sizes were not statistically significant.

Adjusted Analysis. In fully adjusted GEE models, results shown in Table 20, water quality was not associated with child stunting, and the adjusted RR was almost 1.0 for both drinking and non-drinking water sources.

Table 20

Adjusted Association of Water Source Quality and Child Stunting at 18 Months

Water access variable	n	Adjusted RR	95% CI		*p*
			lower	upper	
Primary source of water for drinking is protected	3958	1.01	.93	1.10	.80

Primary source of water for other uses is protected	3940	1.01	.92	1.11	.82

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: HIV status, height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor. RR=relative risk; CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Interaction Between Maternal HIV Status and Water Source Quality on Child LAZ

Results of tests for interaction between maternal HIV status and water source quality on child LAZ are summarized in Table 21. I found a statistically significant interaction between maternal HIV-status and drinking water source quality, the estimated difference-in difference β was -0.189 (95% CI: -0.367, -0.012); $p = 0.04$. Due to this significant interaction, I tested RQ3 stratified by maternal HIV-status.

Table 21

Interaction Between Water Source Quality and Maternal HIV Status on Child LAZ at 18 Months

Variable	n	β	95% CI		*p*
			LL	UL	
Maternal HIV status × Primary source of water for drinking is protected	4018	-.189	-.367	-.012	**.04**
Maternal HIV status × Primary source of water for other uses is protected	4000	.073	-.127	.273	.48

Note. β = Difference-in-difference estimate; CI=confidence interval. Interaction $P<0.10$ was considered as evidence of interaction.

Association of Water Source Quality and Child LAZ Among HIV-Unexposed Children

Among households with HIV-negative mothers, the quality of primary water sources used for drinking as well as for other household purposes was not associated with child linear growth. In the adjusted analysis, results shown in Table 22, the quality of both water sources used for drinking and for other purposes was not

associated with child LAZ.

Table 22

Adjusted Association of Water Source Quality and LAZ at 18 Months Among HIV-Unexposed Children

Water access variable	n	Adjusted mean difference	95% CI		*p*
			LL	UL	
Primary source of water for drinking is protected	3316	-.052	-.120	.018	.15
Primary source of water for other uses is protected	3299	.005	-.074	.083	.91

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor; CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Association of Water Source Quality and Child LAZ among HIV-Exposed Children

In unadjusted analysis, the water quality of primary water source for drinking was significantly associated with improved child linear growth among children born to HIV-infected women. The estimated change in LAZ was 0.180 (95% CI: 0.016, 0.344); $p = 0.03$. However, non-drinking water source quality was not associated with child linear growth.

After fully adjusting for confounders, the association of drinking water source quality and child linear growth among children born to HIV- infected women was no longer statistically significant. Table 23 shows the adjusted effects of water source quality on LAZ among children born to HIV-infected mothers. The estimated adjusted mean difference in child LAZ was 0.045 (95% CI: -0.122, 0.211); $p = 0.60$. There was very weak evidence of an association between non-drinking water source quality and child LAZ; the adjusted mean difference in LAZ was -0.143 (95% CI: -0.306, 0.020);

$p = 0.09$, however the direction of the estimated association went in the opposite direction to what I had hypothesized.

Table 23

Adjusted Association of Water Source Quality and LAZ at 18 Months Among HIV-Exposed Children

Variable	n	Adjusted mean difference	95% CI		*p*
			LL	UL	
Primary source of water for drinking is protected	642	.045	-.122	.211	.60
Primary source of water for other uses is protected	641	-0.143	-0.306	0.020	**.09**

Note. The regression models were adjusted for study related factors: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; Maternal factors: height, MUAC, and employment status; Child factors: birth weight, sex, and prematurity; and household factors: SES, improved floor; CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Effect of Community Water Coverage on Child Linear Growth

To test the hypothesis that child linear growth was associated with community-level water coverage, first I constructed a dataset where I computed the summary exposure and outcome variables per cluster. I computed the following summary measures by cluster: mean LAZ, prevalence of stunting, median of total volume used by households, median of water volume per capita per day, median distance to water source, median one-way walk-time to water source, prevalence of safe drinking water, and summary of other covariates used for adjusting regression models. I then conducted univariable and multivariable linear regression analyses on these cluster-level variables to test the hypotheses.

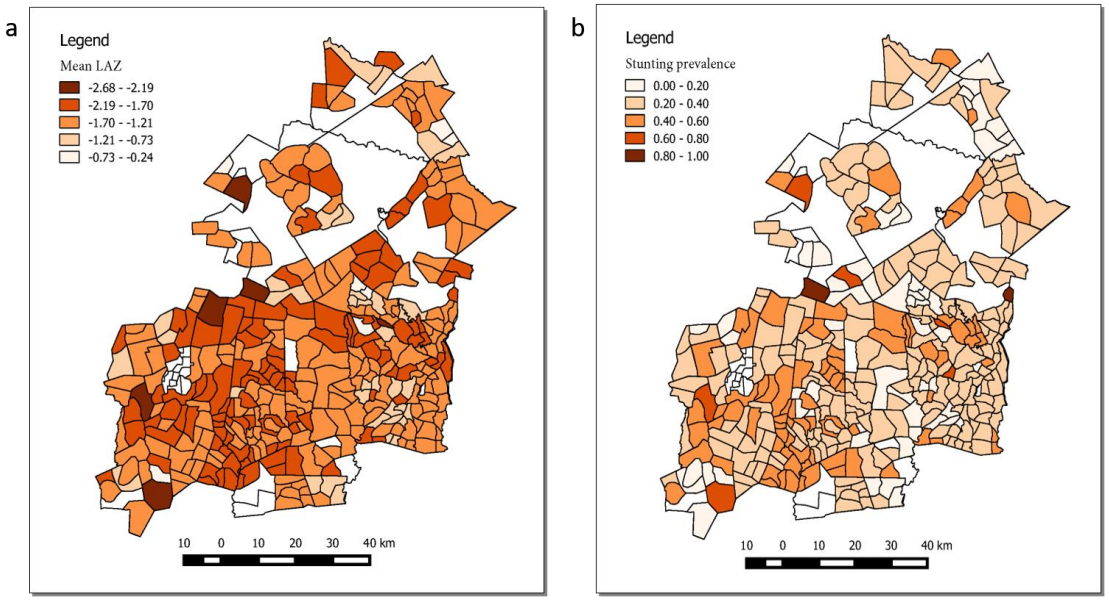
Spatial Distribution of Cluster-Level Child Linear Growth

To visualize the spatial distribution of child linear growth across the study

area, I plotted maps showing the spatial distribution of LAZ and stunting across the study area shown in Figure 17. Mean LAZ for each cluster ranged from -0.24 to -2.68, with most clusters having mean LAZ in the middle two bands: -2.19 to -1.70 and -1.70 to -1.21 (see panel A). The map also shows that there are some clusters with severely retarded child growth, having LAZ down to -2.68 in the western regions of the study area. The average stunting prevalence per cluster ranged from less than 1% to over 90% as shown in Figure 17 panel B. Child stunting was clustered around certain regions within the study area, and it is then important to explore if these same regions had correspondingly poor water coverage, which is explored in the next section.

Figure 17

Maps Showing the Spatial Distribution of Cluster-Level Child Growth



Note. Panels (a) shows mean LAZ by cluster with deeper red for lower LAZ; and (b) shows stunting prevalence by cluster with light color representing lower stunting.

Spatial Distribution of Cluster-Level Water Access

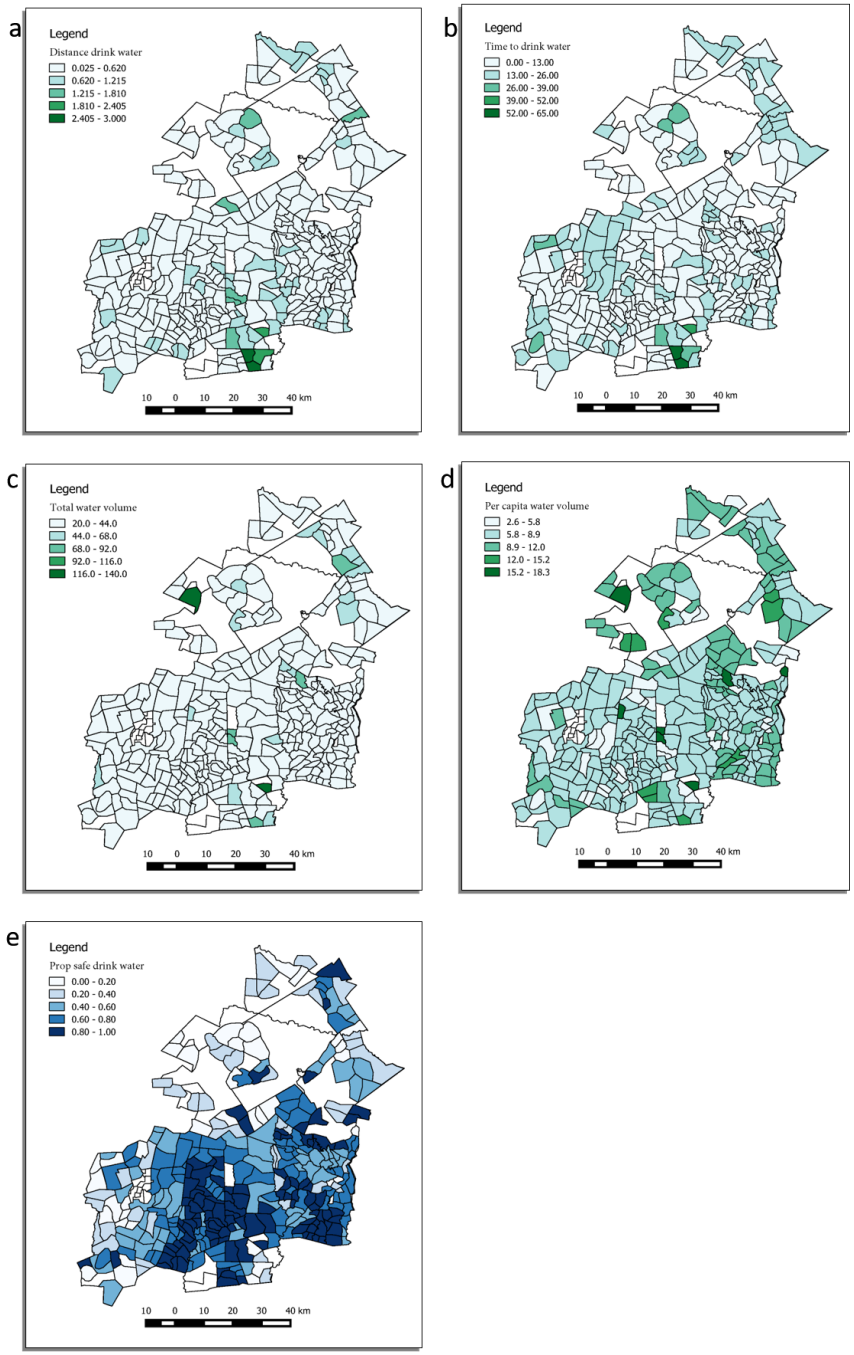
To visualize the spatial distribution of different water access measures across

the study area, I plotted five maps shown in Figure 18 that depicts the variation of cluster-level median total water volume, median per capita water volume, median distance to drinking water source, median time to drinking water source, and proportion of safe drinking water. The first map in panel (A) shows the variation in mean distance to drinking water source by cluster. The map shows that households in most clusters were between 25m to 620m with a few clusters up to 3km. The second map in panel (B) shows the spatial variation of one-way walk-time to the primary drinking water source by cluster. There appears to be a wider spatial variation in walk-times compared to distance, however there is some correlation between the distance and time maps. The third map in panel (C), shows the spatial distribution of median water volume by cluster. The map suggests that the variation in the medians of total water volume of water collected by households per day is wide. This clearly shows that most households collected between 20 to 44 liters and very few households collected larger volumes up to 140 liters. The fourth map panel (D) shows the spatial variation of per capita water volume used by household members per day. There is a wide spatial variation in the cluster median per capita water volume used by household members per day. The map also shows that household in the eastern region of the study area used more water than those on the western regions. This correlates well with the observation from Figure 16 that stunting was less prevalent in the eastern region compared to the western region. The last map panel (E) shows the spatial distribution of safe drinking water by cluster. This map shows that the prevalence of safe drinking water sources is higher in the southern region than the northern. However, this distribution does not correlate well with the observed spatial distribution of LAZ or stunting that means that there may be a weak or no association

between drinking water quality and child growth in the study area.

Figure 18

Maps Showing the Spatial Distribution of Cluster-Level Water Access



Note. Panels (a) median distance to drinking water (in km) by cluster, the deeper green color represents longer distance; (b) median walk-time to drinking water by cluster, here the deeper green color represents longer times; (c) median of total water volume by cluster, here the deeper green color represents more water; (d) median volume of water per capita per day in each cluster, the deeper green color represents more water used; (e) prevalence of safe drinking water by cluster, the deeper blue color represents a higher proportion of households with safe water.

Exploring Cluster-Level Mean LAZ and Cluster-Level Water Coverage

I explored trends between mean cluster-level length-for-age z-score and cluster-level water coverage indicators. I plotted bivariate scatterplots between the cluster-level mean LAZ and various cluster-level water coverage indicators that include: median total water volume, median per capita water volume, median distance to drinking water source, median time to drinking water source, and proportion of safe drinking water. The scatterplots showed a trend generally depicting that: Clusters with higher average water volume had better child growth, with the exception of two outlier clusters with median volume > 100 liters but low mean LAZ. The trend between mean LAZ and median per capita volume was non-linear; mean LAZ and per capita water volume was positively correlated for lower per capita water volumes however progressively poorer growth when per capita volume >15 liters per person per day. The trend between mean LAZ and median distance to drinking water was however counter intuitive – clusters with shorter median distance to drinking water exhibited poorer child growth. I also observed a similar trend with median time to drinking water, however there were a few outlier clusters that could have influenced this observed trend. A more consistent trend was observed in the scatterplot of mean LAZ by proportion with safe drinking water; clusters with higher prevalence of safe drinking water had better child mean LAZ. I performed a similar exploratory analyses to understand the bivariate association between LAZ and water access to non-drinking water and observed similar trends.

Association Between Cluster-Level Mean LAZ and Water Access

Results of unadjusted linear regression models testing the association between cluster-level mean LAZ and cluster-level water access, summarized in Supplementary

Table B20, showed that neither the volume of water nor quality of the water used by households was associated with mean LAZ. However physical accessibility of water sources for both drinking and non-drinking was strongly associated with mean LAZ. The median distance to water sources for both drinking and non-drinking were strongly associated with child linear growth. The mean difference in cluster-level mean LAZ (95% CI) was 0.0001 (0.00001, 0.0003); $p = 0.03$ and 0.002 (0.00001, 0.0004); $p = 0.004$ for each additional meter walked respectively. The median time to water sources for both drinking and non-drinking were associated with child linear growth, mean difference in cluster-level mean LAZ (95% CI) was 0.006 (-0.0004, 0.013); $p = 0.07$ and 0.007 (0.0006, 0.013); $p = 0.03$ for each additional minute walked respectively. Results of unadjusted analyses could be affected by confounding as the point estimates are going in the opposite direction to the hypothesized, they all suggest that clusters that were further from water sources were associated with better child growth.

Results shown in Table 24 are the fully adjusted linear regression models of mean LAZ by cluster-level water access measures. All the community-level water access measures – median distance, median time, and prevalence of safe water were not associated with cluster-level mean LAZ. Results that were statistically significant in unadjusted analyses were no longer significant after accounting for confounding.

Table 24

Adjusted Association Between Cluster-Level Mean LAZ at 18 Months and Cluster-Level Water Access

Water access variable	Adjusted β	95% CI		*p*
		LL	UL	

Median of total volume of water used by households in each cluster (liters)	.001	-.003	.004	.67
Median of per-capita volume of water used by each household member in each cluster (liters)	.004	-.014	.023	.65
Median of distance to drinking water by households in each cluster (m)	.0001	-.0001	.0002	.21
Median of distance to water used for other uses by households in each cluster (m)	.0001	-.00002	.0003	.11
Median of time to drinking water by households in each cluster (minutes)	.004	-.002	.010	.18
Median of time to water used for other uses by households in each cluster (minutes)	.004	-.003	.010	.26
Proportion of households with safe drinking water per cluster (%/10)	-.002	-.019	.015	.80
Proportion of households with safe water used for other uses per cluster (%/10)	.004	-.021	.028	.77

Note. N=209 clusters, β=Difference in mean LAZ, CI=confidence interval; LL=lower limit; UL=upper limit. All regression models were adjusted for the following factors: maternal- mean age, mean height, mean nutrition status (mid-upper arm circumference MUAC), and proportion HIV negative; Child: proportion of female children, prevalence of low birthweight, prevalence of premature; Household: mean SES score, proportion of households with any type latrine, proportion of households practicing open defecation; study arm: WASH and IYCF. Significant associations are in bold face.

Exploring Cluster-Level Proportion Stunted and Cluster-Level Water Coverage

I explored graphically the bivariate associations between prevalence of child stunting and the same five water coverage indicators. Scatterplots showing cluster-level prevalence of child stunting by cluster-level drinking water coverage exhibited a general trend showing that clusters with higher average water volume had lower prevalence of child stunting, however two clusters were outliers, which was consistent with the LAZ analysis. For median per capita water volume, the scatterplot showed that the prevalence of stunting was negatively correlated with median per capita water volume, two clusters were also visible as outliers at the high end of water volumes per capita. The same trend in stunting prevalence is observed with both distance to water source and walk-time to water source. Clusters with higher prevalence of child stunting had lower prevalence of safe water for drinking. I performed a similar

exploration to understand the bivariate association between child stunting and water access to non-drinking water sources and found similar trends.

Association Between Cluster-Level Proportion Stunted and Cluster-Level Water Access

I first conducted unadjusted analyses using fractional logistic regression models to test the association between cluster level prevalence of stunting and cluster-level water access, results shown in Supplementary Table B21. Volume of water used by households was the only measure not associated with child stunting, difference in proportion stunted (95% CI) -0.001 (-0.002, 0.001); $p = 0.30$. Median per-capita volume of water used by household members in each cluster was weakly associated with reduction in prevalence of stunting, difference in proportion stunted (95% CI) -0.007 (-0.015, 0.001); $p = 0.09$. The median of distance to non-drinking water by households had a stronger unadjusted association with stunting prevalence compared to distance to drinking water, similar effect size - difference in proportion stunted (95% CI) -0.0001 (-0.0001, -0.0000); $p = 0.04$ and $p = 0.005$ respectively. The median of walk-time to non-drinking water by households also had a stronger unadjusted association with stunting prevalence compared to walk-time to drinking water, difference in proportion stunted (95% CI) -0.003 (-0.006, -0.0004); $p = 0.025$ and -0.003 (-0.005, 0.000); $p = 0.05$ respectively. The prevalence of safe water for drinking was however more strongly associated with the prevalence of child stunting compared to non-drinking water, the difference in proportion stunted was (95% CI) -0.008 (-0.014, -0.001); $p = 0.028$ compared to -0.010 (-0.020, -0.0001); $p = 0.048$.

I then conducted adjusted analyses using fractional logistic regression models to test the association between cluster level prevalence of stunting and cluster-level

water access adjusting for confounding variables, results are summarized in Table 25.

In adjusted analyses, only two water access measures remained statistically significant. Median distance to water used for other uses by households in each cluster was associated with the proportion of stunted children in that cluster, the estimated adjusted difference in the proportion of stunted children was -0.0001 (95% CI: -0.0001, 0.0000); $p = 0.06$. Median time to water to drinking water by households in each cluster was associated with the proportion of stunted children in that cluster, the estimated adjusted difference in the proportion of stunted children was -0.0021 (95% CI: -0.0047, 0.0004); $p = 0.10$. All other community-level water coverage measures were not associated with child stunting at 18 months of age.

Table 25

Adjusted Association of Mean Stunting at 18 Months with Median Distance to Primary Water Sources Used by Household

Water access variable	Adjusted β	95% CI		*p*
		LL	UL	
Median of total volume of water used by households in each cluster (liters)	-0.001	-0.002	.001	.42
Median of per-capita volume of water used by each household member in each cluster (liters)	-0.004	-0.011	.004	.33
Median of distance to drinking water by households in each cluster (m)	-0.00003	-0.0001	.00002	.23
Median of distance to water used for other uses by households in each cluster (m)	-0.0001	-0.0001	.000	**.06**
Median of time to drinking water by households in each cluster (minutes)	-0.002	-0.005	.0004	.10
Median of time to water used for other uses by households in each cluster (minutes)	-0.002	-0.005	.001	.11
Proportion of households with safe drinking water per cluster (%/10)	-0.003	-0.010	.004	.47
Proportion of households with safe water used for other uses per cluster (%/10)	-0.003	-0.013	.007	.57

Note. N=209 clusters, β=Difference in proportion stunted, CI=confidence interval; LL=lower limit; UL=upper limit; All regression models were adjusted for the following factors:- maternal: mean age, mean height, mean nutrition status (mid-upper arm circumference MUAC), and proportion HIV negative; Child: proportion of female children, prevalence of low birthweight, prevalence of premature; Household: mean SES score, proportion of households with any type latrine, proportion of

households practicing open defecation; study arm: WASH and IYCF. Significant associations (p<0.10) are in bold face.

Summary

In this section, I presented the purpose and research questions for the study and the hypotheses that were tested. I discussed the statistical methodology used for analyzing the data and described data transformations used. I described an outline of how the 4,036 participants who met inclusion criteria for this study were drawn from the secondary dataset. I then presented their summary characteristics, including socio-demographic and study settings, nutritional status as well as water access characteristics. Of note is that one third of children in the study were stunted; mean LAZ = -1.57 and 16% of the mothers were HIV-infected. I presented analysis of each research question in turn, presenting descriptive statistics, unadjusted and adjusted regression models. I tested each exposure-outcome pair for effect modification by maternal HIV exposure and presented stratified analyses when there was a significant interaction. I found that maternal HIV exposure was an important effect modifier some of the associations. Finally, I tested the effects of community-level water coverage on child linear growth; first constructing maps showing the spatial distribution of exposures and outcomes, then running regression models to estimate effect sizes. In section 4 I interpreted these findings in relation to the theoretical framework, conceptual models, and compare the findings with those found in the literature. I then make recommendations and present my conclusions.

Section 4: Application to Professional Practice and Implications for Social Change

Introduction

The purpose of this quantitative prospective cohort study was to explore the effects of access to water on child linear growth and stunting at 18 months of age among children born in two rural districts in Zimbabwean. This study was aimed at assessing if access to water was an important determinant of child growth and stunting in a rural setting in LMICs. The study was conducted using secondary data from the SHINE study, a cluster-randomized clinical trial that was conducted in two rural districts in Zimbabwe. The current study included a large cohort of 4, 036 mother/infant dyads who had matched water access and child linear growth measurements. Study participants were drawn from contiguous clusters across two districts in central Zimbabwe and balanced across the four study arms of the original trial (Humphrey et al., 2018; Prendergast et al., 2019; SHINE Team, 2015).

In this section, I summarize the key findings of this study and interpret them in relation to the theoretical framework, conceptual models, and the literature. I therefore compare the study findings with those found in literature and discuss how these findings add to the body of knowledge. I conclude by making recommendations based on the findings, discussing the implications for social change of these results, then I present my conclusions.

Summary of Key Findings

The context of the study setting and participants was well described. Women were on average in their late 20s and had medium height. They had mostly attained secondary education, a vast majority of women were unemployed, and most were married. Further, they received good nutritional diet; 40% met the minimum dietary

needs and had normal hemoglobin levels. These are among some of the known maternal risk-factors for child linear growth (Amugsi et al., 2019; Prado, et al., 2019). The dominant religion was “apostolic faith” comprising 45% of the women, which was reported as a strong risk factor for maternal and child health underutilization in Africa (Ha Salama et al., 2014). HIV prevalence was high among the women, as 16% were HIV-positive; however, HIV incidence among children was very low with only ~ 3% of the children being positive at 18 months of age. A high proportion (22%) of children were born premature and 8% had birthweight < 2,500g. The prevalence of child stunting in the study districts was high, 33.3% of the children were stunted at 18 months with a mean LAZ of -1.57.

WaSH coverage was low in the study setting; only just over a third, 36% of the women had an improved latrine at the household and there were virtually no piped water systems in this population with ~1% having water piped into their homes. The predominant sources of water were boreholes used by two thirds of households for drinking. Unprotected surface and ground water was used by 65% of households for non-drinking purposes. During the implementation of the original study, there were 3-4% more women in the WASH and WASH+IYCF arms of the study compared to those in the SOC and IYCF arms.

The results from regression analyses of the association between child LAZ or stunting and water access showed that water volume per capita per day was positively associated with improved child linear growth. The effect was more pronounced among children born to mothers who were HIV-infected but not for those born to HIV-negative mothers. Distance to water for non-drinking purposes was weakly associated with LAZ, but one-way walk time to any water source was not associated

with child linear growth. Water quality was not associated with child linear growth; however, on stratifying results by maternal HIV-status there was very weak evidence of an association between non-drinking water quality and child LAZ.

Using spatial analysis, I revealed some patterns in the occurrence of stunting and water access the study area. Mean LAZ among the clusters ranged from -0.24 to -2.68, with most clusters having mean LAZ in the middle two bands. Average stunting prevalence per cluster ranged between 1% and 90%. Median distance to primary water sources in most clusters was between 25m to 620m, with a few clusters up to 3km away. One-way walk time to the primary drinking water source varied more by cluster when compared to distance; however, there was strong correlation between two measures. Total water collected by households per day varied widely between clusters from 20 to 44 liters; very few households collected larger volumes up to 140 liters per day, which could indicate other uses such as agricultural uses or laundry. However, all the community-level water access measures: median distance, median time, and prevalence of safe water were not associated with mean LAZ. The only association that I found with community water-coverage was weak between median distance to non-drinking water and the proportion of stunted children.

Interpretation

Finding 1: Effect of Water Quantity on Child Linear Growth

In this study I showed that the volume of water used by households was clustered in multiples of 20 liters and can be attributed this to the nature of containers used for collecting water because the most common types of containers used measured 20 liters. I found that gross volume of water used by a household was not associated with child LAZ and stunting at 18 months of age. These finding were

consistent with earlier observations that quantities of water used by rural households do not vary much unless the water source was either very close to the home (on plot) in that case they used a large volume or very far (Cairncross & Feachem, 1993; Howard et al., 2020). This is a water plateau of indifference; hence, child linear growth may also not vary with total water volume used (Howard et al., 2020).

Water volume used by each household member per day was positively associated with improved child linear growth. This was an important finding in that even with small average water volume per person found in this study, this was still able to have a significant effect on child linear growth. The average per capita water in this population was only median 7.1 liters (IQR: 5, 11.7) even though the UN recommends 50 to100 liters per person per day (UN & WHO, 2010). The findings are consistent with those reported on the effect of water volume on infectious morbidity where ~6 and 11 liters per person per day were needed to see a difference in reported diarrhea cases in Kenya and Ghana respectively (De Buck et al., 2015). The findings however differed with studies that implemented WaSH interventions with childhood stunting as the outcome and reported mixed findings for both morbidity and growth (Cumming & Cairncross, 2016). Some studies have reported positive benefits of WaSH improving growth or reducing morbidity, and other studies reviewed found no benefit of WaSH on either child growth or morbidity.

I also found a significant interaction between maternal HIV-status during pregnancy and water volume used by the household on child linear growth. The effect of water volume on child linear growth was different between HIV-exposed compared to HIV-unexposed children. Among children born to HIV-negative women, there was no evidence of association between child growth and water volume. However,

children born to women living with HIV had better liner growth when they had access to higher volumes of water in the household when compared to children born to HIV-negative women during pregnancy. It is known that HIV-exposed children are more likely to catch infectious morbidity even when they themselves are not infected with HIV (Koyanagi, et al., 2011; Evans, et al., 2021). Higher water volume promotes better hygienic practices, which could have resulted in lower exposure to infectious morbidity. Nutritional interventions have also been shown to be more efficacious in children exposed to HIV compared to those unexposed; an intervention combining WaSH and small quantity lipid-based nutrient supplement (sqLNS) was more effective in improving neurodevelopment scores among HIV-exposed children compared nutrition or WaSH interventions given separately (Chandna et al., 2020) Hence, access to higher quantities of water among these HIV-exposed children may also have resulted in better linear growth.

Finding 2: Effect of Household Distance from Water Source on Child Linear Growth

I was able to examine the effect of distance to water according to the intended use of the water—drinking and non-drinking household purposes. The ability to separate these effects was an added strength of this study. However, I did not find any association between child linear growth and distance to drinking water source. This finding agrees with results that distance to water source was an independent predictor for underweight but not stunting in both Tanzania and Ghana (Abubakar et al., 2012). This results however contradict findings that distance walked to fetch freshwater uniquely predicted childhood stunting and underweight in Demographic Health Survey analyses from SSA countries (Pickering & Davies, 2012).

I also found very weak evidence of an association between distance to non-drinking water source and child linear growth. This points to a possible link between child linear growth and hygiene. A similar association between WaSH and child growth was observed but only in observational and not experimental studies, and attributed the association to the effects of unobserved and unmeasured confounding (Stewart et al., 2019).

The median distance to a primary water source in this study was 400 meters (IQR: 100, 1,000) and less than 1% of households had piped water connection into their homes. This distance is too far from the home to affect volumes of water used and hence the health of children (Cairncross & Feachem, 1993). A water plateau is when the volume of water used declined sharply over the first few meters from the home, levelled out as distance increased up to 1,000 meters, thereafter the volumes declined further with distance. The lack of association between distance to water sources and child linear growth in this study was therefore consistent with their findings.

Finding 3: Effect of Walk-time to Primary Water Source on Child Linear Growth

Time taken to walk to a water source is strongly correlated with distance to the source but also reflects the effort imposed by the terrain or other obstacles in the path. In the analysis I found that one-way walk-time to a primary water source for drinking was only weakly associated with child LAZ and stunting in unadjusted analyses and child stunting was not associated with walk-time to non-drinking water sources in both unadjusted and adjusted analyses. The lack of effect of one-way walk-time to water and child linear growth was consistent with the previous finding with distance

to water as reported by Howard et al. (2020). Walk-time to water sources is usually trusted to capture better the effort to fetch water than just Euclidean distance. It is also believed that most households poorly estimate distance as compared to time hence should be a better estimate of water access. Maternal HIV-status in pregnancy did not modify the association between one-way walk-time to water and child linear growth. This was not expected as I had anticipated that water fetching would be more difficult for HIV-infected women.

Finding 4: Effect of Primary Water Source Quality on Child Linear Growth

The quality of water used by households in the study was classified according to the degree of protection of the water source. Chemical analysis of water quality was not reported in the original study. Generally water quality in rural SSA is not monitored for quality (Smiley, 2017). I found the use of both safe and unprotected water high in the study; In Malawi and other SSA countries, households make complex decisions to use both clean and unsafe water in their homes (Smiley, 2017). Non-drinking water is then collected from open water sources while drinking water collected from protected sources. I found that child linear growth was not associated with the quality of water used by households for either drinking or other purposes, I did not find any differences in the impact of drinking or non-drinking water on child linear growth. The level of water pollution found in rural Zimbabwean water sources may not be as very high to affect child health in a similar was to that found in the cities where sewerage and industrial effluent are sometimes discharged into rivers. However, I found a strong interaction between maternal HIV status and water quality suggesting that the effect of water quality on child linear growth is modified by infant exposure to maternal HIV and I conducted stratified analyses to test the hypothesis

among HIV-exposed and HIV-unexposed children separately.

Among HIV-unexposed children, water quality was not associated with either LAZ or stunting. However, among HIV-exposed children I found a significant association between drinking water quality and LAZ, but only in unadjusted analysis where the mean change in LAZ was 0.18 standard deviations higher among children born to HIV-infected women and had access to clean water compared to those without. The effect was however not significant after adjusting for confounders. I did not expect this since HIV-exposed children have been shown to lag in many growth and development domains when compared to HIV-unexposed children (Rosala-Hallas et al., 2019; Ntozini, et al., 2020), and are more amenable to interventions (Chandna, et al., 2020), making them respond more to minor improvements in water quality than HIV-unexposed children. In this study however, I did not find any evidence supporting this and possibly due to lower levels of pollution in rural water sources to make a significant difference in linear growth.

Finding 5: Effect of Community Water Coverage on Child Linear Growth

In the study, I derived community water coverage from the clusters and used regression analyses to test the hypothesis. Spatial analysis revealed some uneven distribution of water access and child growth. I found significant associations between child growth and several community-level water coverage indicators, but only in unadjusted analyses. In adjusted models, LAZ was no longer associated with any of the water coverage indicators. Stunting was however weakly associated with some coverage indicators, distance and time to both drinking and non-drinking water sources. These had very small but positive impacts on child growth such that a small increase in distance was associated with a corresponding increase in the proportion of

stunted children or a reduction in mean LAZ. Overall, community-level water access was weakly associated with child liner growth that is in contrast with results reported in Mali where found strong evidence of an association between community sanitation coverage and child growth was reported (Harris et al., 2017). In Mali they were looking at the full WaSH coverage whilst in this study I only examined water access and water alone may not have a strong enough effect on child linear growth.

Most evidence for WaSH interventions have been from observational studies and very few experimental studies, especially provision of water. According to the Provision of water is a human right, hence any researcher designing an experimental study to examine the benefits of access to water must tread the fine path balancing between the human rights, ethical obligation of providing water to the needy and the scientific rigor of experimental studies (UN & WHO, 2010). Observational studies are prone to both measured and unmeasured confounding biases (Creswell, 2009; Delgado-Rodriguez & Llorca, 2004; Hulley et al., 2007), hence a step-wedge design might offer the best way to evaluate the effects of water access while satisfying both needs (Hemming et al., 2015). A step-wedge design is a quasi-experimental design that allows a staggered rollout of a water access improvement program over time while evaluating health outcomes before and after receiving the intervention to compare effectiveness.

Limitations

The study was observational, that is water access was not assigned by the researcher but only observed during the conduct of the trial, hence the findings are prone to confounding with other unmeasured biases (Creswell, 2009; Delgado-Rodriguez & Llorca, 2004; Hulley et al., 2007). Results can only be associational and

not taken as causal.

This research study was limited to households that participated in the original trial meeting the inclusion criteria for that study. The study is limited to mothers and children who resided in the two study districts during 2012 to 2017.

Most of the limitations of secondary data apply to this study since it was collected by others for their own specific purposes (Glass, 1976; Smith et al., 2011). Some of the limitations associated with using these data to conduct new research (Cheng and Philips, 2014), include: (a) Data may be incomplete, some key variables could be missing in the accessible dataset to enable a more complete and proper interpretation of results obtained. For example, GPS coordinates of the households would have provided a means to plot spatial distribution of the sample in the districts to get more accurate maps, however these were not available –removed as a measure of de-identifying participants to protect their identity and confidentiality; (b) The data may be too old if data were collected and the situation has since changed so much that data may be less representative of the current situation in the study area. The economic situation in Zimbabwe is fast changing and the conditions that obtained in 2012-2015 when baseline assessments were carried out and 2017-2018 when endline interviews were done might be quite different now; (c) Some variables may have been measured on a different scale than what would be ideal to answer the current research questions. Here community water access was never an objective of the original trial however I created proxy variables to estimate community coverage that may not be the true coverage in those communities; (d) Variable definitions could have changed over time during data collection in the trial. For example, access to sanitation changed during the conduct of the trial from the baseline since toilets were provided as part of

the trial interventions. Socio-economic status may also have changed during the conduct of the trial because the trial provided some inputs such as cell phones and solar chargers that are normally factored into assessments of SES; (e) the study sample may not be representative of the larger population from that they were collected such that results will not be generalizable to that population. Here the sample of women and children included in the trial may be different to the population of the study districts such that the results may not be generalizable to the two districts; and (f) the dataset may have some other biases reflecting the views of the researchers who conducted the primary study. For example, water access was assessed by maternal interview that could be prone to recall biases.

Recommendations

This study was designed to analyze secondary data to examine associations between water coverage in rural LMIC countries and child linear growth. I found some important associations that need a study specifically designed to examine these hypotheses.

Water access has been defined in many different ways in different communities, hence it is important to be specific which definition is used for any study. Several definitions can be used in order to triangulate the effects. In this study I used several definitions to describe water access that includes distance, time and quality. Each captures a different aspect of water access that together gives a full description of water access.

HIV modified the effect of water access on child growth. In the study I found improvements in water access had better child linear growth among children born to HIV-infected mothers, whether they themselves are infected or just exposed to HIV.

In populations where HIV is a major public health challenge it is important check intervention effects stratified by HIV exposure. Interventions focusing on this vulnerable population of children should include provision of clean water.

Community coverage is an important indicator of the degree of access a target population has, for example to a public health intervention. Community water coverage is hard to quantify accurately from survey data not designed for spatial data collection. In the study, I derived useful summary data that can be used for further analysis to understand population risk. I found some interesting associations between LAZ and stunting with the derived community water coverage indicators.

The use of spatial analysis also helps reveal subtle patterns in the data and should be used studies or programs examining water access and health. Plotting spatial maps showing distribution of public health outcomes of interest is an important way to visualize potential hotspots of public health challenges. In the study, I showed the spatial distribution of child stunting, water quantity and quality, as well as time taken to fetch water. With additional spatial data for the study area, one can also examine other geographical features and climatic conditions that may affect water access such as the terrain, drainage, rainfall pattern, among others to understand how they influence water access.

Rural water access alone is not enough to prevent child stunting. In this study I did not find any significant associations between water access and child linear growth. I found water was between 100-400m from homes which is still too far and many household still used less than recommended water quantity of 15-20 liters per day. To have positive impact on health and growth of children, water sources may have to be brought into the homes to lessen the burden of water fetching. The provision of water

to rural populations as per the current recommended in the SDGs “within 1,000m” is not enough to impact child linear growth as results of this study have shown.

Water quality should be accurately assessed. Depending on the degree of protection of the water source may not capture the true quality of water used in rural homes. I did not find a strong association between water quality and child growth maybe because water may have been contaminated on the way to the home. In the absence of a chemical water quality assessment it is hard to assess the true effects of water quality on child linear growth.

Child stunting is a complex public health problem that has remained stubborn for decades and requires multiple levels of interventions. Focusing on one type of intervention at a time may not explain or solve the problem of stunting. Combining multiple interventions know to have an effect of child health may have synergistic effects.

Other health conditions or comorbidities affecting children may complicate their response to interventions. For example I found a strong interaction between water access and exposure to HIV through infected mothers. Other chronic diseases may also have similar effects on child stunting modifying the effects of water access on linear growth.

Implications for Social Change

Stunting remains such a public health problem that it is now a target for reduction included among the UN targets since the last decade; initially among the MDGs and now in the SDGs (UN, n.d.-a; UN, n.d.-b). Finding solutions to the child-stunting problem has remained a high priority over the last decade. Finding a solution to this problem would have an important public health impact and a saving on

resources being expended in search for solutions.

Results of this research could potentially inform local authorities and beyond with information on the expected impact of their current and future water access interventions in similar settings. The results could be used to guide decision making on the optimal provision of water services in similar settings; For example, this research showed that distance or walk-time to water sources will not impact child health if the distance is beyond a certain threshold. This could be used to inform the design of future rural water projects if child health is the outcome.

Covering whole communities has also remained a sought after target for WaSH interventions wishing to make greater impact in those communities. This research showed that community water coverage is important but not enough to impact child linear growth alone. Other complementary interventions may be required to complement water access improvements to impact child health.

This study showed that investing in water quality interventions in the study district was not able to impact child linear growth except for some vulnerable populations. The need for targeted public health interventions; –precision public health, continues to be promoted as the most effective way to address current public health challenges. Identifying vulnerable population sub-groups that could benefit more from each intervention is very important in order to optimize resources. In this research, I identified a potential vulnerable population that could benefit from targeted water access interventions with greater impact. HIV exposure was only one of many potential exposures that make sub-populations more vulnerable; hence, more research is needed to identify such populations and design effective interventions for them.

Conclusion

In this study, the effect of water access on child linear growth was assessed using data from a cluster-randomized study conducted in two rural districts in Zimbabwe. Over 4,000 mother-infant dyads were included in the analyses. The study was conducted is a population with high prevalence of stunting and maternal HIV. Analysis of effects of various water access measures on LAZ and stunting showed that volume of water per capita per day, improved water quality, and community-level coverage were all important determinants of child linear growth in the study. Furthermore, the growth of children born to HIV-infected mothers showed greater response to water access compared to HIV-unexposed children who showed no impact. Distance or walk-time to a primary water source showed no impact on child linear growth that was attributed to having water sources located too far from the households to change consumption levels sufficiently to affect child health. Results of this study showed the importance of identifying vulnerable populations that could be targeted by public health interventions to have better impact. The study also showed that, for maximum impact, it is important to design interventions that target individuals as well as community-level factors that determine population health. Finally, this study suggested that further research with more appropriate study designs to answer questions on water access and child growth.

References

Abubakar, A., Uriyo, J., Msuya, S. E., Swai, M., & Stray-Pedersen, B. (2012). Prevalence and risk factors for poor nutritional status among children in the Kilimanjaro region of Tanzania. *International Journal of Environmental Research and Public Health, 9*(10), 3506–3518. https://doi.org/10.3390/ijerph9103506

Adair, L. S., Fall, C. H., Osmond, C., Stein, A. D., Martorell, R., Ramirez-Zea, M., Sachdev H. S., Dahly D. L., Bas I., Norris S. A., Micklesfield L., Hallal P., Victora C. G., & COHORTS group. (2013). Associations of linear growth and relative weight gain during early life with adult health and human capital in countries of low and middle income: Findings from five birth cohort studies. *The Lancet, 382*(9891), 525–534. https://doi: 10.1016/S0140-6736(13)60103-8.

Aiga, H., & Umenai, T. (2003). Standardisation of the definition of access to safe water. *The Lancet, 361*(9375), 2156. https://doi.org/10.1016/S0140-6736(03)13703-8

American College Health Association. (2018). Healthy Campus 2020: Ecological Model. *American College Health Association, Hanover*. https://acha.org/HealthyCampus/Implement/Ecological_Model/HealthyCampus/Ecological_Model.aspx

Amugsi, D. A., Dimbuene, Z. T., & Kimani-Murage, E. W. (2019). Factors associated with normal linear growth among pre-school children living in better-off households: A multi-country analysis of nationally representative data. https://doi.org/10.1101/797845

Asiamah, N., Mensah, H. K., & Oteng-Abayie, E. (2017). General, target, and accessible population: Demystifying the concepts for effective sampling. *The Qualitative Report, 22*(6), 1607–1621. https://nsuworks.nova.edu/tqr/vol22/iss6/9

Barry, A., & Honoré, H. (2009). Everyday theory: A practical application of the ecological perspective. *American Journal of Health Education, 40*(6), 368–372. https://doi.org/10.1080/19325037.2009.10599116

Bartram, J., & Cairncross, S. (2010). Hygiene, sanitation, and water: Forgotten foundations of health. *PLoS Medicine, 7*(11), e1000367. https://doi.org/10.1371/journal.pmed.1000367

Baruth, M., & Wilcox, S. (2013). Multiple behavior change among church members taking part in the faith, activity, and nutrition program. *Journal of Nutrition Education and Behavior, 45*(5), 428–434. https://doi.org/10.1016/j.jneb.2013.03.002

Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., De Onis, M., Ezzati, M., Mathers C., Rivera J., & Maternal and Child Undernutrition Study Group. (2008). Maternal and child undernutrition: global and regional exposures and health consequences. *The lancet, 371*(9608), 243–260. https://doi.org/10.1016/S0140-6736(07)61690-0

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., De Onis, M., Ezzati M., Grantham-McGregor S., Katz J., Martorell R., Uauy R., P, & Maternal and Child Nutrition Study Group. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet, 382*(9890), 427–451. https://doi.org/10.1016/S0140-

6736(13)60937-X

Bronfenbrenner, U. (1989). Ecological systems theory. In R. Vasta, *Annals of Child Development: Vol. 6* (pp. 187–249). Jessica Kingsley Publishers.

Bronfenbrenner, U. (1994). Ecological models of human development. *International Encyclopedia of Education, 3*(2), 37–43. http://edfa2402resources.yolasite.com/resources/Ecological%20Models%20of%20Human%20Development.pdf

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. *Handbook of Child Psychology*. https://doi.org/10.1002/9780470147658.chpsy0114

Bruni, M., & Spuhler, D. (2020). Water source protection. Sustainable sanitation and water management toolbox - SSWM.info. https://sswm.info/arctic-wash/module-4-technology/further-resources-water-sources/water-source-protection

Cairncross, S., & Feachem, R. (1993). *Environmental health engineering in the tropics: an introductory text* (No. Ed. 2). John Wiley & Sons.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

Centers for Disease Control and Prevention. (2015). Colorectal Cancer Control Program (CRCCP). https://www.cdc.gov/cancer/crccp/sem.htm

Centers for Disease Control and Prevention. (2018). The social-ecological model: A framework for prevention. https://www.cdc.gov/violenceprevention/about/social-ecologicalmodel.html

Chandna, J., Ntozini, R., Evans, C., Kandawasvika, G., Chasekwa, B., Majo, F.D., Mutasa, K., Tavengwa, N.V., Mutasa, B., Mbuya, M.N., & Moulton, L.H., (2020). Effects of improved complementary feeding and improved water, sanitation and hygiene on early child development among HIV-exposed children: substudy of a cluster randomised trial in rural Zimbabwe. *BMJ Global Health, 5*(1), e001718. https://doi.org/10.1136/bmjgh-2019-001718

Cheng, H. G., & Phillips, M. R. (2014). Secondary analysis of existing data: opportunities and implementation. *Shanghai Archives of Psychiatry, 26*(6), 371. https://dx.doi.org/10.11919/j.issn.1002-0829.214171

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association, 74:*368, 829–836. https://doi.org/10.1080/01621459.1979.10481038

Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*(403), 596–610. https://doi.org/10.1080/01621459.1988.10478639

Corraini, P., Olsen, M., Pedersen, L., Dekkers, O. M., & Vandenbroucke, J. P. (2019). Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators [Corrigendum]. *Clinical Epidemiology, 11,* 245. https://dx.doi.org/10.2147/CLEP.S129728

Creswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. *Thousand Oaks, Calif. Sage publications.*

Cumming, O., & Cairncross, S. (2016). Can water, sanitation and hygiene help eliminate stunting? Current evidence and policy implications. *Maternal &*

Child Nutrition, 12(S1), 91–105. https://doi.org/10.1111/mcn.12258

Dangour, A. D., Watson, L., Cumming, O., Boisson, S., Che, Y., & Velleman, Y. (2013). The effect of interventions to improve water quality and supply, provide sanitation and promote handwashing with soap on physical growth in children. *Cochrane Syst Rev, 8,* CD009382. https://doi.org/10.1002/14651858.CD009382.pub2

De Buck, E., Borra, V., De Weerdt, E., Veegaete, A. V., & Vandekerckhove, P. (2015). A systematic review of the amount of water per person per day needed to prevent morbidity and mortality in (post-) disaster settings. *PloS One, 10*(5), e0126395. https://doi.org/10.1371/journal.pone.0126395

de Onis, M., & Branca, F. (2016). Childhood stunting: a global perspective. *Maternal & Child Nutrition, 12*(S1), 12–26. https://doi.org/10.1111/mcn.12231

de Onis, M., Blössner, M., & Borghi, E. (2012). Prevalence and trends of stunting among pre-school children, 1990–2020. *Public Health Nutrition, 15*(01), 142–148. https://doi.org/10.1017/S1368980011001315

Delgado-Rodriguez, M., & Llorca, J. (2004). Bias. *Journal of Epidemiology & Community Health, 58*(8), 635–641. https://doi.org/10.1136/jech.2003.008466

Dewey, K. G., & Begum, K. (2011). Long-term consequences of stunting in early life. *Maternal & Child Nutrition, 7*(s3), 5–18. https://doi.org/10.1111/j.1740-8709.2011.00349.x

Diersing, N., & Nancy, F. (2009). Water quality: Frequently asked questions. *Florida Brooks National Marine Sanctuary, Key West, FL.* https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=1458406

Dudovskiy, J. (2018). Research Methodology: Regression Analyses. *Necessary knowledge to conduct a business research.* https://research-methodology.net/research-methods/quantitative-research/regression-analysis/

Esrey, S. A., Potash, J. B., Roberts, L., & Shiff, C. (1991). Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma. *Bulletin of the World Health Organization, 69*(5), 609–621. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2393264/

Evans, C., Chasekwa, B., Ntozini, R., Majo, F. D., Mutasa, K., Tavengwa, N., Mutasa B., Mbuya M. N. N., Smith, L. E., Stoltzfus R. J., Moulton, L. H., Humphrey, J. H., Prendergast, A. J., & SHINE Trial Team. (2021). Mortality, HIV transmission and growth in children exposed to HIV in rural Zimbabwe. *Clinical Infectious Diseases, 72*(4), 586–594. https://dx.doi.org/10.1093/cid/ciaa076

Frankfort-Nachmias, C., Nachmias, D., & DeWaard, J. (2015). Research Methods in the Social Sciences (8th Edition). *Worth Publishers. New York.*

Fry, L. M., Cowden, J. R., Watkins Jr, D. W., Clasen, T., & Mihelcic, J. R. (2010). Quantifying health improvements from water quantity enhancement: An engineering perspective applied to rainwater harvesting in West Africa. *Environmental science & technology, 44*(24), 9535–9541. https://doi.org/10.1021/es100798j

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5(10), 3-8.* https://doi.org/10.3102/0013189X005010003

Global Nutrition Profile. (2014). 2014 Nutrition Country Profile: Economics and

Demography Child Anthropometry Adolescent and Adult Nutrition Status Zimbabwe. *The International Food Policy Research Institute. Washington DC.* http://globalnutritionreport.org/files/2014/12/gnr14_cp_zimbabwe.pdf

Government unveils new child health card, vaccine. (2012, June 30). *Herald House, Harare.* https://www.herald.co.zw/government-unveils-new-child-health-card-vaccine/

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Ha, W., Salama, P., Gwavuya, S., & Kanjala, C. (2014). Is religion the forgotten variable in maternal and child health? Evidence from Zimbabwe. *Social Science & Medicine, 118,* 80–88. https://doi.org/10.1016/j.socscimed.2014.07.066

Harris, M., Alzua, M. L., Osbert, N., & Pickering, A. (2017). Community-level sanitation coverage more strongly associated with child growth and household drinking water quality than access to a private toilet in rural Mali. *Environmental Science & Technology, 51*(12), 7219–7227. https://doi.org/10.1021/acs.est.7b00178

Hemming, K., Haines, T. P., Chilton, P. J., Girling, A. J., & Lilford, R. J. (2015). The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ, 350*. https://doi.org/10.1136/bmj.h391

Howard, G., Bartram, J., Williams, A., Overbo, A., Geere, J. A., & WHO. (2020). Domestic water quantity, service level and health, second edition. *World*

Health Organization, Geneva. https://www.who.int/publications/i/item/9789240015241

Huang, F. L. (2019). Alternatives to logistic regression models in experimental studies. *The Journal of Experimental Education, 1-16.* https://doi.org/10.1080/00220973.2019.1699769

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G., & Newman, T. B. (2007). Designing clinical research. *Williams and Wilkins, a Walters Kluwer business, 503 Walnut Street;* pp. 251–65.

Humphrey, J. H. (2009). Child undernutrition, tropical enteropathy, toilets, and handwashing. *The Lancet, 374*(9694), 1032–1035. https://doi.org/10.1016/S0140-6736(09)60950-8

Humphrey, J. H., Mbuya, M. N., Ntozini, R., Moulton, L. H., Stoltzfus, R. J., Tavengwa, N. V., Mutasa, K., Majo, F., Mutasa, B., Mangwadu, G., Chasokela, C. M., Chigumira, A., Chasekwa, B., Smith, L. E., Tielsch, J. M., Jones, A. D., Manges, A. R., Maluccio, J. A., Prendergast, A. J., & SHINE Trial Team. (2019). Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on child stunting and anaemia in rural Zimbabwe: a cluster-randomised trial. *The Lancet Global Health, 7*(1), e132–e147. https://doi.org/10.1016/S2214-109X(18)30374-7

IBM Corp. (Released 2017). Statistical Package for the Social Science (SPSS) version 25. *IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM.*

Koyanagi, A., Humphrey, J. H., Ntozini, R., Nathoo, K., Moulton, L. H., Iliff, P., Mutasa, K., Ruff, A., Ward, B., & ZVITAMBO Study Group. (2011).

Morbidity among human immunodeficiency virus-exposed but uninfected, human immunodeficiency virus-infected, and human immunodeficiency virus-unexposed infants in Zimbabwe before availability of highly active antiretroviral therapy. *The Pediatric Infectious Disease Journal, 30*(1), 45–51. https://doi.org/10.1097/INF.0b013e3181ecbf7e

Kumar, S., Quinn, S. C., Kim, K. H., Musa, D., Hilyard, K. M., & Freimuth, V. S. (2012). The Social Ecological Model as a Framework for Determinants of 2009 H1N1 Influenza Vaccine Uptake in the US. *Health Education & Behavior: The Official Publication of the Society for Public Health Education, 39*(2), 229–243. https://doi.org/10.1177/1090198111415105

Lehmann, J., Jones, D., Wilson-Jones, M., Wheeler, C., Callister, F., Steele, R., Gupta, P., Randriarimalala, E., Shiferaw, B., & Chanda, G. (June 2016). Caught short: How a lack of toilets and clean water contributes to malnutrition. *WaterAid.* https://www.wateraid.org/in/sites/g/files/jkxoof336/files/caught-short-how-a-lack-of-toilets-and-clean-water-contributes-to-malnutrition.pdf

Little, R. J., & Rubin, D. B. (2014). Statistical analysis with missing data (Vol. 333). *John Wiley & Sons. Hoboken, New Jersey.*

Liu, X. (2015). Methods and applications of longitudinal data analysis. *Elsevier*. https://doi.org/10.1016/B978-0-12-801342-7.00009-5

Luby, S. P., Rahman, M., Arnold, B. F., Unicomb, L., Ashraf, S., Winch, P. J., Stewart, C. P., Begum, F., Hussain, F., Benjamin-Chung, J., Leontsini, E., Naser, A.M., Parvez, S. M., Hubbard, A. E., Lin, A., Nizame, F. A., Jannat, K., Ercumen, A., Ram, P. K., ... , Colford Jr, J. M. (2018). Effects of water

quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial. *The Lancet Global Health*, *6*(3), e302–e315. https://doi.org/10.1016/S2214-109X(17)30490-4

McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An ecological perspective on health promotion programs. *Health Education Quarterly*, *15*(4), 351–377. https://doi.org/10.1177/109019818801500401

Misra, S. (2012). Randomized double blind placebo control studies, the "Gold Standard" in intervention based studies. *Indian Journal of Sexually Transmitted Diseases*, *33*(2), 131. pp. 147–59. https://dx.doi.org/10.4103/2589-0557.102130

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821). *John Wiley & Sons.*

Murray, G. D. (1998). Missing Data in Clinical Trials. *Wiley StatsRef: Statistics Reference Online.* https://doi.org/10.1002/9781118445112.stat04946

Murtaugh, P. A. (2014). In defense of P values. *Ecology*, *95*(3), 611–617. https://doi.org/10.1890/13-0590.1

Ntozini, R., Chandna, J., Evans, C., Chasekwa, B., Majo, F. D., Kandawasvika, G., Tavengwa, N. V., Mutasa, B., Mutasa, K., Moulton, L. H., Humphrey, J. H., Gladstone, M. J., Prendergast, A. J., & SHINE Trial Team. (2020). Early child development in children who are HIV-exposed uninfected compared to children who are HIV-unexposed: observational sub-study of a cluster-randomized trial in rural Zimbabwe. *Journal of the International AIDS Society*, *23*(5), e25456. https://doi.org/10.1002/jia2.25456

Ntozini, R., Humphrey, J., Prendergast, A., & Moulton, L. H. (2022, October 12). The SHINE Trial - Zimbabwe. *Open Science Framework.* https://doi.org/10.17605/OSF.IO/W93HY

Ntozini, R., Marks, S. J., Mangwadu, G., Mbuya, M. N. N., Gerema, G., Mutasa, B., Julian, T. R., Schwab, K. J., Humphrey, J. H., Zungu, L. I., & SHINE Trial Team. (2015). Using Geographic Information Systems and Spatial Analysis Methods to Assess Household Water Access and Sanitation Coverage in the SHINE Trial. *Clinical Infectious Diseases, 61,* S716–S725. https://doi.org/10.1093/cid/civ847

Null, C., Stewart,C. P., Pickering, A. J., Dentz, H. N., Arnold, B. F., Arnold, C. D., Benjamin-Chung, J., Clasen, T., Dewey, K G., Fernald, L. C. H., Hubbard, A. E., Kariger, P., Lin, A., Luby, S. P., Mertens, A., Njenga, S. M., Nyambane, G., Ram, P. K., & Colford Jr, J. M. (2018). Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial. *The Lancet Global Health, 6*(3), e316–e329. https://doi.org/10.1016/S2214-109X(18)30005-6

Nygren, B. L., O'Reilly, C. E., Rajasingham, A., Omore, R., Ombok, M., Awuor, A. O., Jaron, P., Moke, F., Vulule, J., Laserson, K., Farag, T. H., Nasrin, D., Nataro, J. P., Kotloff, K. L., Levine, M. M., Derado, G., Ayers, T. L., Lash, R. R., Breiman, R. F., & Mintz, E. D. (2016). The relationship between distance to water source and moderate-to-severe diarrhea in the global enterics multi-center study in Kenya, 2008–2011. *The American Journal of Tropical Medicine and Hygiene*, *94*(5), 1143–1149. https://doi.org/10.4269/ajtmh.15-0393

Overbo, A., Williams, A. R., Evans, B., Hunter, P. R., & Bartram, J. (2016). On-plot drinking water supplies and health: A systematic review. *International Journal of Hygiene and Environmental Health, 219*(4–5), 317–330. https://doi.org/10.1016/j.ijheh.2016.04.008

Panjwani, A., & Heidkamp, R. (2017). Complementary Feeding Interventions Have a Small but Significant Impact on Linear and Ponderal Growth of Children in Low-and Middle-Income Countries: A Systematic Review and Meta-Analysis. *The Journal of Nutrition, jn243857*. https://doi.org/10.3945/jn.116.243857

Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics, 11*(6), 619–632. https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1

Pearson, A. L. (2016). Comparison of methods to estimate water access: a pilot study of a GPS-based approach in low resource settings. *InternationalJjournal of Health Geographics, 15(1), 33.* https://doi.org/10.1186/s12942-016-0062-8

Peletz, R., Simuyandi, M., Sarenje, K., Baisley, K., Kelly, P., Filteau, S., & Clasen, T. (2011). Drinking water quality, feeding practices, and diarrhea among children under 2 years of HIV-positive mothers in peri-urban Zambia. *The American Journal of Tropical Medicine and Hygiene, 85*(2), 318–326. https://doi.org/10.4269/ajtmh.2011.11-0140

Pickering, A. J., & Davis, J. (2012). Freshwater availability and water fetching distance affect child health in sub-Saharan Africa. *Environmental Science & Technology, 46*(4), 2391–2397. https://doi.org/10.1021/es203177v

Prado, E. L., Jimenez, E. Y., Vosti, S., Stewart, R., Stewart, C. P., Somé, J., Pulakka, A., Ouédraogo, J. B., Okronipa, H., Ocansey, E., Oaks, B., Maleta, K., Lartey, A., Kortekangas, E., Hess, S. Y., Brown, K., Bendabenda, J., Ashorn, U., Ashorn, P., ... , Dewey, K. (2019). Path analyses of risk factors for linear growth faltering in four prospective cohorts of young children in Ghana, Malawi and Burkina Faso. *BMJ Global Health*, *4*(1), e001155. https://dx.doi.org/10.1136/bmjgh-2018-001155

Prendergast, A. J. & Humphrey, J. H. (2014) The stunting syndrome in developing countries. *Paediatrics and International Child Health*, *34*:4, 250–265. https://doi.org/10.1179/2046905514Y.0000000158

Prendergast, A. J. , Chasekwa, B., Evans, C., Mutasa, K., Mbuya, M. N. N., Stoltzfus, R. J., Smith, L. E., Majo, F. D., Tavengwa, N. V., Mutasa, B., Mangwadu, G. T., Chasokela, C. M., Chigumira, A., Moulton, L. H., Ntozini, R., Humphrey, J. H., & SHINE Trial Team. (2019). Independent and combined effects of improved water, sanitation, and hygiene, and improved complementary feeding, on stunting and anaemia among HIV-exposed children in rural Zimbabwe: a cluster-randomised controlled trial. *The Lancet Child & Adolescent Health*, *3*(2), 77–90. https://doi.org/10.1016/S2352-4642(18)30340-7

Prüss-Ustün, A., Bartram, J., Clasen, T., Colford Jr, J. M., Cumming, O., Curtis, V., Bonjour, S., Dangour, A. D., De France, J., Fewtrell, L., Freeman, M. C., Gordon, B., Hunter, P. R., Johnston, R. B., Mathers, C., Mäusezahl, D., Medlicott, K., Neira, M., Stocks, M., ... , Cairncross, S. (2014). Burden of disease from inadequate water, sanitation and hygiene in low-and middle-

income settings: a retrospective analysis of data from 145 countries. *Tropical Medicine & International Health*, *19*(8), 894–905. https://doi.org/10.1111/tmi.12329

Ramsey, F., & Schafer, D. (2012). The statistical sleuth: a course in methods of data analysis. *Cengage Learning. Brooks Cole Publishing Company, CA.*

Rasella, D. (2013). Impact of the Water for All Program (PAT) on childhood morbidity and mortality from diarrhea in the Bahia State, Brazil. *Cadernos de Saúde Pública*, *29*(1), 40–50. https://scielosp.org/scielo.php?pid=S0102-311X2013000500006&script=sci_arttext

Rosala-Hallas, A., Bartlett, J. W., & Filteau, S. (2017). Growth of HIV-exposed uninfected, compared with HIV-unexposed, Zambian children: a longitudinal analysis from infancy to school age. *BMC Pediatrics*, *17*(1), 80. https://doi.org/10.1186/s12887-017-0828-6

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys (*Wiley series in probability and statistics). Applied Probability and Statistics. New York.*

Said-Mohamed, R., Micklesfield, L. K., Pettifor, J. M., & Norris, S. A. (2015). Has the prevalence of stunting in South African children changed in 40 years? A systematic review. *BMC Public Health*, *15*(1), 534. https://doi.org/10.1186/s12889-015-1844-9

Schmidt, C. W. (2014). Beyond malnutrition: the role of sanitation in stunted growth. *Environmental Health Perspectives*, *122*(11), A298. https://doi.org/10.1289/ehp.122-A298

SHINE Trial Team. (2015). The Sanitation Hygiene Infant Nutrition Efficacy (SHINE) trial: rationale, design, and methods. *Clinical Infectious Diseases,*

61(suppl 7), S685–S702. https://doi.org/10.1093/cid/civ844

Smiley, S. L. (2017). Quality matters: incorporating water quality into water access monitoring in rural Malawi. *Water International*, *42*(5), 585–598. https://doi.org/10.1080/02508060.2017.1344818

Smith, A. K., Ayanian, J. Z., Covinsky, K. E., Landon, B. E., McCarthy, E. P., Wee C. C., & Steinman, M. A. (2011). Conducting high-value secondary dataset analysis: An introductory guide and resources. *Journal of General Internal Medicine*, *26*(8), 920–929. https://doi.org/10.1007/s11606-010-1621-5

Spears, D. (2013). How much international variation in child height can sanitation explain? World Bank policy research working paper, (6351). *The World Bank Sustainable Development Network Water and Sanitation Program.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2212559

Stark, C. M., Devine, C. M., & Dollahite, J. S. (2017). Characteristics associated with the application of an ecological approach to preventing childhood obesity. *Public Health Nutrition*, *20*(1), 174–182. https://doi.org/10.1017/S1368980016001798

StataCorp. (2019). Stata 16 Base Reference Manual: Fractional outcome regression. *College Station, TX: Stata Press.* https://www.stata.com/features/overview/fractional-outcome-models/

Stein, A. D., Thompson, A. M., & Waters, A. (2005). Childhood growth and chronic disease: evidence from countries undergoing the nutrition transition. *Maternal & Child Nutrition*, *1*(3), 177–184. https://doi.org/10.1111/j.1740-8709.2005.00021.x

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G.,

Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ, 338*, b2393. https://doi.org/10.1136/bmj.b2393

Stewart, C. P., Iannotti, L., Dewey, K. G., Michaelsen, K. F., & Onyango, A. W. (2013). Contextualising complementary feeding in a broader framework for stunting prevention. *Maternal & Child Nutrition, 9*(S2), 27–45. https://doi.org/10.1111/mcn.12088

Tamason, C. C., Bessias, S., Villada, A., Tulsiani, S. M., Ensink, J. H., Gurley, E. S., & Mackie Jensen, P. K. (2016). Measuring domestic water use: a systematic review of methodologies that measure unmetered water use in low-income settings. *Tropical Medicine & International Health, 21*(11), 1389–1402. https://doi.org/10.1111/tmi.12769

U.S. Department of Health and Human Services. (n.d). Phase I report: recommendations for the framework and format of healthy people 2020. http://www.healthypeople.gov/2010/hp2020/advisory/phasei/summary.htm

UN & WHO. (2010). The human right to water and sanitation: media briefing - UN-Water Decade Programme on Advocacy and Communication and Water Supply and Sanitation Collaborative Council. *Water Supply and Sanitation Collaborative Council (WSSCC). Geneva.* https://www.un.org/waterforlifedecade/pdf/human_right_to_water_and_sanitation_media_brief.pdf

UNICEF & WHO. (1990). Strategy for improved nutrition of children and women in developing countries. *A Unicef Policy Review. UNICEF New York.* http://repository.forcedmigration.org/pdf/?pid=fmo:3066

UNICEF, WHO, & International Bank for Reconstruction and Development/The World Bank. (2020). Levels and trends in child malnutrition: Key Findings of the 2020 Edition of the Joint Child Malnutrition Estimates. *Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.* https://data.unicef.org/resources/joint-child-malnutrition-estimates-interactive-dashboard-2020/

UNICEF, WHO, & World Bank Group. (2018). Levels and trends in child malnutrition: Key findings of the 2018 Edition of the Joint Child Malnutrition Estimates. *UNICEF New York.* https://data.unicef.org/resources/levels-and-trends-in-child-malnutrition-2018/

UNICEF. (2009). Communication for Development (C4D): Realizing Strategic Shifts and Accelerating Results for Children. *Policy and Practice. UNICEF New York.* http://www.unicef.org/cbsc/files/C4D-Position-Paper.doc

United Nations. (n.d.-a). Millennium Development Goals. *UN, Geneva.* http://www.un.org/millenniumgoals/

United Nations. (n.d.-b). Sustainable Development Goals Knowledge Platform. *UN, Geneva.* https://sustainabledevelopment.un.org/sdgs

Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., Sachdev, H. S., & Maternal and Child Undernutrition Study Group. (2008). Maternal and child undernutrition: consequences for adult health and human capital. *The Lancet, 371*(9609), 340–357. https://doi.org/10.1016/S0140-6736(07)61692-4

Victora, C. G., de Onis, M., Hallal, P. C., Blössner, M., & Shrimpton, R. (2010). Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics, peds–2009.* https://doi.org/10.1542/peds.2009-1519

Vollmer, S., Harttgen, K., Subramanyam, M. A., Finlay, J., Klasen, S., & Subramanian, S. V. (2014). Association between economic growth and early childhood undernutrition: evidence from 121 Demographic and Health Surveys from 36 low-income and middle-income countries. *The lancet Global Health, 2*(4), e225–e234. https://doi.org/10.1016/S2214-109X(14)70025-7

Wacholder, S. (1986). Binomial regression in GLIM: estimating risk ratios and risk differences. *American journal of epidemiology, 123*(1), 174–184. https://doi.org/10.1093/oxfordjournals.aje.a114212

Wang, M. (2014). Generalized estimating equations in longitudinal data analysis: a review and recent developments. *Advances in Statistics, 2014.* https://dx.doi.org/10.1155/2014/303728

Wedderburn, C. J., Evans, C., Yeung, S., Gibb, D. M., Donald, K. A., & Prendergast, A. J. (2019). Growth and neurodevelopment of HIV-exposed uninfected children: a conceptual framework. *Current HIV/AIDS Reports, 16*(6), 501–513. https://doi.org/10.1007/s11904-019-00459-0

WHO Multicentre Growth Reference Study Group. (2006). WHO Child Growth Standards: Methods and development: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age. *World Health Organization, Geneva.* http://www.who.int/childgrowth/publications/technical_report_pub/en/

WHO & UNICEF Joint Monitoring Program (JMP). (2014). Zimbabwe: estimates on the use of water sources and sanitation facilities (1980 - 2015). *World Health Organization, Geneva.* http://www.wssinfo.org/documents/?tx_displaycontroller[type]=country_files

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. *MIT press.* http://www.mmf.lnu.edu.ua/images/stories/depart/mee/wooldridge_cross_section_and_panel_data.pdf

World Health Organization. (2016a). Child growth standards: Length/height-for-age. *World Health Organization, Geneva.* http://www.who.int/childgrowth/standards/height_for_age/en/

World Health Organization. (2016b). Global Database on Child Growth and Malnutrition. *World Health Organization, Geneva.* http://www.who.int/nutgrowthdb/database/countries/zwe/en/

World Health Organization. (2023). The UNICEF/WHO/ World Bank Joint Child Malnutrition Estimates (JME) group released new data for 2021. *World Health Organization, Geneva.* https://www.who.int/news/item/06-05-2021-the-unicef-who-wb-joint-child-malnutrition-estimates-group-released-new-data-for-2021

World Health Organization. (7 February 2018). Drinking-water. *World Health Organization, Geneva.* http://www.who.int/en/news-room/fact-sheets/detail/drinking-water

World Health Organization and UNICEF (WHO/UNICEF). (2014). Progress on Drinking-Water and Sanitation: 2014 update. *World Health Organization, Geneva.* https://www.unicef.org/publications/files/JMP_report_2014_webEng.pdf

Zimbabwe National Statistics Agency & ICF International. (2016). Zimbabwe Demographic and Health Survey 2015: Final Report. *Zimbabwe Population*

Census Office. Harare. Rockville, Maryland, USA: https://dhsprogram.com/pubs/pdf/FR322/FR322.pdf

Zimbabwe National Statistics Agency. (2013a). Census 2012: National Report. *Zimbabwe Population Census Office. Harare.* http://www.zimstat.co.zw/sites/default/files/img/National_Report.pdf

Zimbabwe National Statistics Agency (ZIMSTAT). (2013b). Census 2012: Provincial Report Midlands. *Zimbabwe Population Census Office. Harare.* http://www.zimstat.co.zw/sites/default/files/img/publications/Census/CensusResults2012/Midlands.pdf

Appendix A: Permission Letter

ZVITAMBO

Institute for Maternal and Child Health Research

Investigator/Study Staff Site Agreement

Sanitation Hygiene Infant Efficacy Project (SHINE)

I agree to assume the responsibility as Student Investigator_____ *(insert role: "Principal Investigator", "co-investigator"; "student investigator" whose work on the project is associated with an academic degree objective; or "study staff", meaning personnel who will interface with participants and/or their identifiable private information, such as research coordinator(s), interviewer (s), or data manager(s))* for the study listed above. I understand that this responsibility includes all of the following commitments:

1. I will protect the rights and welfare of all study participants.
2. I will follow the IRB approved research plan.
3. I will not institute any changes for which IRB review is required, to the research plan or any other study documentation without prior IRB approval.
4. I will comply with IRB policies, and with the federal, state, international, or local laws applicable to the site of the research.

I do [ ] do not [x] (check one) have a financial conflict of interest with this study. If you check "do", please disclose your conflict to the Principal Investigator on the project, who should report it and discuss further necessary actions with the IRB.

Robert Ntozini

Print Name (Student Investigator)

Signature _____ Date 18/09/2019

For Zvitambo Institute for Maternal and Child Health Research Professor Jean H. Humphrey, Sc.D_____ Print Name (Principal Investigator and Director)

Signature _____ Date 19/09/2019

No. 16 Lauchlan Avenue, Meyrick Park, Mabelreign, Harare, Zimbabwe
Tel +263-242-306028/306056//306259/ Fax: 311069. E-mail: administrator@zvitambo.co.zw

Appendix B: Supplementary Results

Table B1

Unadjusted Association of Water Volume Used by Household and LAZ at 18 Months

Water Access Variable	n	Unadjusted mean difference	95% CI		*p*
			LL	UL	
Total water volume fetched to the home in 24hrs (liters)	3441	.001	.0002	.002	**.01**
Log_{10}(Total water volume)	3441	.143	.032	.253	**.01**
Water volume per person per day (liters)	3328	.003	-.001	.008	.17
Log_{10}(Water volume per person per day)	3328	.106	-.006	.218	**.06**

Note. CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Table B2

Partially Adjusted Association of Water Volume Used by Household and LAZ at 18 Months

Water Access Variable	n	Partially adjusted mean difference	95% CI		*p*
			LL	UL	
Total water volume fetched to the home in 24hrs (liters)	3440	.001	.0003	.002	**.007**
Log_{10}(Total water volume)	3440	.142	.029	.254	**.01**
Water volume per person per day (liters)	3327	.004	-.001	.008	.13
Log_{10}(Water volume per person per day)	3327	.110	.0004	.220	**.05**

Note. All regression models were adjusted for: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Table B3

Unadjusted Association of Water Volume Used by Household and Child Stunting at 18 Months

Water Access Variable	n	Unadjusted RR	95% CI		*p*
			LL	UL	
Total water volume fetched to the home in 24hrs (liters)	3441	.999	.997	1.00	**.05**
Log_{10}(Total water volume)	3441	.92	.80	1.07	.27
Water volume per person per day (liters)	3328	.998	.99	1.00	.53
Log_{10}(Water volume per person per day)	3328	.96	.82	1.12	.56

Note. CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Table B4

Partially Adjusted Association of Water Volume Used by Household and Child Stunting at 18 Months

Water Access Variable	n	Partially adjusted RR	95% CI		*p*
			LL	UL	
Total water volume fetched to the home in 24hrs (liters)	3440	.999	.997	1.00	**.04**
Log_{10}(Total water volume)	3440	.93	.809	1.06	.27
Water volume per person per day (liters)	3327	.997	.99	1.00	.39
Log_{10}(Water volume per person per day)	3327	.95	.82	1.10	.47

Note. All models were adjusted for the following covariates: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment; CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B5

Interaction Between Water Volume Measures and Maternal HIV Status on LAZ at 18 Months

Interaction term	n	β	95% CI		*p*
			LL	UL	
Maternal HIV status × Total water volume fetched to the home in 24hrs (liters)	3441	-.003	-.007	.001	**.09**
Maternal HIV status × Log_{10}(Total water volume)	3441	.062	-.458	.334	.76
Maternal HIV status × Water volume per person per day (liters)	3328	.011	-.028	.005	.18
Maternal HIV status × Log_{10}(Water volume per person per day)	3328	-.030	-.380	.320	.87

Note. β = Difference-in-difference estimate; CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face; Interaction $p<0.10$ was considered as evidence of interaction.

Table B6

Unadjusted Association of Water Volume Used by Household and LAZ at 18 Months Among HIV-Unexposed Children

Water Access Variable	n	Unadjusted Mean difference	95% CI		*p*
			LL	UL	
Total water volume fetched to the home in 24hrs (liters)	2892	.0007	.00003	.0013	**.04**
Log_{10}(Total water volume)	2892	.116	-.003	.235	**.06**
Water volume per person per day (liters)	2793	.002	-.002	.007	.35
Log_{10}(Water volume per person per day)	2793	.106	-.011	.223	**.07**

Note. The regression models were adjusted for study arm only – WASH and IYCF; CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B7

Unadjusted Association of Water Volume Used by Household and LAZ at 18 Months Among HIV-Exposed Children

Water access variable	n	Unadjusted mean difference	95% CI		*p*
			LL	UL	
Total water volume fetched to the home in 24hrs (liters)	549	.004	.0004	.008	**.03**
Log_{10}(Total water volume)	549	.185	-.173	.551	.31
Water volume per person per day (liters)	535	.0134	-.002	.029	**.09**
Log_{10}(Water volume per person per day)	535	.135	-.185	.454	.41

Note. The regression models were adjusted for study arm only – WASH and IYCF; CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Table B8

Unadjusted Association of Distance to Primary Water Source and LAZ at 18 Months

Water access variable	n	Unadjusted mean difference	95% CI		*p*
			LL	UL	
Distance to primary water source for drinking (m)	4014	-.000	-.0001	.000	.64
Log_{10}(Distance to primary water source for drinking)	4014	-.010	-.036	.016	.47
Distance to primary water source for other uses (m)	4003	.000	-.00004	.0001	.82

Log_{10}(Distance to primary water source for other uses)	4003	-.023	-.048	.003	.09

Note. CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B9

Partially Adjusted Association of Distance to Primary Water Source and LAZ at 18 Months

Water access variable	n	Partially adjusted mean difference	95% CI		*p*
			LL	UL	
Distance to primary water source for drinking (m)	4011	-.000	-.000	.000	.91
Log_{10}(Distance to primary water source for drinking)	4011	-.003	-.028	.023	.83
Distance to primary water source for other uses (m)	4000	.00002	-.00002	.0001	.44
Log_{10}(Distance to primary water source for other uses)	4000	-.015	-.040	.011	.26

Note. All regression models were adjusted for: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment. CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B10

Unadjusted Association of Distance to Primary Water Source and Child Stunting at 18 Months

Water access variable	n	Unadjusted RR	95% CI		*p*
			lower	upper	
Distance to primary water source for drinking (m)	4014	1.00	1.00	1.00	.26
Log_{10}(Distance to primary water source for drinking)	4014	1.01	.980	1.05	.48
Distance to primary water source for other uses (m)	4003	1.00	.99997	1.00	.46
Log_{10}(Distance to primary water source for other uses)	4003	1.02	.98	1.05	.31

Note. CI=confidence interval; LL=lower limit; UL=upper limit; Significant associations are in bold face.

Table B11

Partially Adjusted Association of Distance to Primary Water Source and Child Stunting at 18 Months

Water access variable	n	Partially adjusted Relative risk	95% CI		*p*
			Lower	Upper	
Distance to primary water source for drinking (m)	4011	1.00	1.00	1.00	.41
Log_{10}(Distance to primary water source for drinking)	4011	1.00	.97	1.03	.91
Distance to primary water source for other uses (m)	4000	1.00	. 99996	1.00	.84
Log_{10}(Distance to primary water source for other uses)	4000	1.01	.98	1.04	.60

Note. All regression models were adjusted for: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment. CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B12

Unadjusted Association of Walk Time to Primary Water Source LAZ at 18 Months

Water access variable	n	Unadjusted mean difference	95% CI		*p*
			LL	UL	
One-way walk-time to primary water source for drinking (mins)	4012	-.002	-.005	-.000	.05
Log_{10}(One-way walk-time to primary water source for drinking)	4012	-.041	-.091	.009	.11
One-way walk-time to primary water source for other uses (mins)	4002	-.0003	-.002	.002	.76
Log_{10}(One-way walk-time to primary water source for other uses)	4002	-.036	-.089	.016	.18

Note. CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B13

Partially Adjusted Association of Walk Time to Primary Water Source and LAZ at 18 Months

Water access variable	n	Partially adjusted mean difference	95% CI		*p*
			LL	UL	

One-way walk-time to primary water source for drinking (mins)	4009	-.002	-.004	.001	.14
Log_{10}(One-way walk-time to primary water source for drinking)	4009	-.016	-.067	.034	.53
One-way walk-time to primary water source for other uses (mins)	3999	.0005	-.002	.003	.61
Log_{10}(One-way walk-time to primary water source for other uses)	3999	-.005	-.057	.046	.85

Note. All regression models were adjusted for: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment. CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B14

Unadjusted Association of Walk Time to Primary Water Source and Child Stunting at 18 Months

Water access variable	n	Unadjusted RR	95% CI		*p*
			LL	UL	
One-way walk-time to primary water source for drinking (mins)	4012	1.00	1.00	1.01	.07
Log_{10}(One-way walk-time to primary water source for drinking)	4012	1.04	.97	1.11	.29
One-way walk-time to primary water source for other uses (mins)	4002	1.00	.999	1.00	.40
Log_{10}(One-way walk-time to primary water source for other uses)	4002	1.04	.97	1.11	.30

Note. CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B15

Partially Adjusted Association of Walk Time to Primary Water Source and Child Stunting at 18 Months

Water access variable	n	Partially adjusted RR	95% CI		*p*
			LL	UL	
One-way walk-time to primary water source for drinking (mins)	4009	.999	1.00	1.00	.21
Log_{10}(One-way walk-time to primary water source for drinking)	4009	1.00	.94	1.07	.95
One-way walk-time to primary water source for other uses (mins)	3999	1.00	.998	1.00	.84
Log_{10}(One-way walk-time to primary water source for other uses)	3999	.9998	.94	1.07	.996

Note. All regression models were adjusted for: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment. CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B16

Unadjusted Association of Water Quality and LAZ at 18 Months

Water access variable	n	Unadjusted mean difference	95% CI		*p*
			LL	UL	
Primary source of water for drinking is protected	4018	.04	-.04	.12	.32
Primary source of water for other uses is protected	4000	.02	-.06	.09	.65

Note. CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B17

Partially Adjusted Association of Water Quality and LAZ at 18 Months

Water access variable	n	Partially adjusted mean difference	95% CI		*p*
			LL	UL	
Primary source of water for drinking is protected	4015	.04	-.03	.11	.27
Primary source of water for other uses is protected	3997	.04	-.04	.11	.35

Note. All regression models were adjusted for: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment. CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B18

Unadjusted Association of Water Quality and Child Stunting at 18 Months

Water access variable	n	Unadjusted RR	95% CI		*p*
			LL	UL	
Primary source of water for drinking is protected	4018	.95	.86	1.05	.29
Primary source of water for other uses is protected	4000	.98	.89	1.08	.69

Note. RR = relative risk; CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B19

Partially Adjusted Association of Water Quality and Child Stunting at 18 Months

Water access variable	n	Partially adjusted RR	95% CI		*p*
			LL	UL	
Primary source of water for drinking is protected	4015	.94	.86	1.04	.22
Primary source of water for other uses is protected	3997	.97	.88	1.06	.45

Note. All regression models were adjusted for: study arm – WASH and IYCF, data collector who took the measurement, and calendar period of assessment. RR = relative risk; CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face.

Table B20

Unadjusted Association Between Cluster-Level Mean LAZ at 18 Months and Cluster-Level Water Access

Water access variable	Unadjusted β	95% CI		*p*
		LL	UL	
Median of total volume of water used by households in each cluster (liters)	.0003	-.002	.004	.88
Median of per-capita volume of water used by each household member in each cluster (liters)	.012	-.009	.030	.31
Median of distance to drinking water by households in each cluster (m)	.0001	.00001	.0003	**.03**
Median of distance to water used for other uses by households in each cluster (m)	.0002	.0001	.0004	**.004**
Median of time to drinking water by households in each cluster (mins)	.001	-0.0004	.013	**.07**
Median of time to water used for other uses by households in each cluster (mins)	.000	0.001	.013	**.03**
Proportion of households with safe drinking water per cluster (%/10)	.065	-0.010	.023	.43
Proportion of households with safe water used for other uses per cluster (%/10)	.013	-0.010	.039	.25

Note. N = 209 clusters, β = Difference in mean LAZ, CI = confidence interval; LL = lower limit; UL = upper limit; Significant associations are in bold face. All regression models were partially adjusted for study arms only: WASH and IYCF.

Table B21

Association Between Mean Stunting at 18 Months and Median Distance to Primary Water Sources Used by Household, Partially Adjusted for Trial Study Arms Only

Water access variable	Unadjusted β	95% CI		*p*
		LL	UL	
Median of total volume of water used by households in each cluster (liters)	-.001	-.002	.001	.30
Median of per-capita volume of water used by each household member in each cluster (liters)	-.007	-.015	.001	.09
Median of distance to drinking water by households in each cluster (m)	-.0001	-.0001	-.000	.04
Median of distance to water used for other uses by households in each cluster (m)	-.0001	-.0001	-.000	.005
Median of time to drinking water by households in each cluster (mins)	-.003	-.005	.000	.05
Median of time to water used for other uses by households in each cluster (mins)	-.003	-.006	-.0004	.025
Proportion of households with safe drinking water per cluster (%/10)	-.008	-.014	-.001	.028
Proportion of households with safe water used for other uses per cluster (%/10)	-.010	-.020	-.0001	.048

Note. N = 209 clusters, β = Difference in proportion stunted, CI = confidence interval; LL = lower limit; UL = upper limit. All regression models were partially adjusted for trial study arms only: WASH and IYCF. Significant associations are in bold face.