Summer 2017

Contributing factors to child stunting in Guatemala: A systems analysis focused on enteric disease transmission and mycotoxin exposure

Lee Emerson Voth-Gaeddert

Department: Civil, Architectural and Environmental Engineering

CONTRIBUTING FACTORS TO CHILD STUNTING IN GUATEMALA: A SYSTEMS ANALYSIS FOCUSED ON ENTERIC DISEASE TRANSMISSION AND MYCOTOXIN EXPOSURE

by

LEE EMERSON VOTH-GAEDDERT

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

CIVIL ENGINEERING

2017

Approved by
Daniel Oerther, Advisor
Robert Holmes
V.A. Samaranayake
Mahelet Fikru
Dev Niyogi

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following five articles which have been submitted for publication, or will be submitted for publication as follows and have been formatted in the style used by each journal:

Paper I: Pages 20-38 have been published by Institute of Electrical and Electronics Engineers Global Humanitarian Technologies Conference Proceedings.

Paper II: Pages 39-61 are intended for submission to the Environmental Health Perspectives Journal.

Paper III: Pages 62-79 are intended for submission to the International Journal of Epidemiology.

Paper IV: Pages 80-99 are intended for submission to the Environmental Health Perspectives Journal.

Paper V: Pages 100-117 have been submitted to the Environmental Science and Technology Journal.

ABSTRACT

Child stunting – low height-for-age – is a United Nation's indicator for chronic malnutrition that has been linked to both acute and chronic health problems. Data from Guatemala suggests for children under five years of age, 49% are classified as stunted. This dissertation tests the following hypotheses, among children in Guatemala 1) environmental enteric dysfunction (EED) is correlated with height-for-age, 2) aflatoxin B (AFB) exposure is correlated with height-for-age, and 3) AFB exposure is correlated with EED. A network analysis was conducted on data from the US Agency for International Development (USAID) collected in 2012 to identify trends in a height-for-age model and an EED model. These results were then combined with a literature review, field observations, and informal interviews to hypothesize two structural equation models (SEM). Additionally, a third SEM was hypothesized for the AFB exposure model. The models were tested with data collected by the San Vicente Health Center in Totonicapán in October 2016 and February 2017. Finally, five geographic specific SEMs were built with the USAID 2012 data and tested with USAID 2013 data. Results of the hypotheses include 1) mixed findings on a correlation between EED and child height-for-age, 2) a confirmed correlation between AFB exposure and child height-for-age, and 3) no correlation between AFB exposure and EED. Furthermore, improved prenatal health and improved sanitary child play areas were correlated with child height-for-age. For the EED model improved water treatment was correlated with reduced EED. Finally, improved maize purchase habits, post-harvest practices, and maize storage were correlated with a decrease in AFB symptoms. Field practitioners and policy makers must account for local and regional suitability for interventions and policies on child health.

ACKNOWLEDGEMENTS

For God, family, and country, in that order. To the glorification of God and His Son, Jesus Christ. To my supportive family, who, without their support and prayers, none of this would be possible. To the country of the United States of America and the Republic of Guatemala, it was a pleasure and honor to serve the people of both countries. Finally, to my adviser, Dr. Daniel B. Oerther, I am forever blessed and grateful for the opportunities you have provided me. Thank you to all.

TABLE OF CONTENTS

APPENDICES

LIST OF ILLUSTRATIONS

LIST OF TABLES

SECTION

1. INTRODUCTION

Low height-for-age, or stunting, is a critical public health indicator, and preventing stunting has been recognized as a global health priority by the United Nations members through the ratification of the Sustainable Development Goals. Goal 2.2 states,

> "By 2030, end all forms of malnutrition, including achieving, by 2025, the internationally agreed targets on stunting and wasting in children under 5 years of age, and address the nutritional needs of adolescent girls, pregnant and lactating women and older persons" [1].

Stunting is defined by the World Health Organization (WHO) as,

> "a height-for-age ratio less than two standard deviations below the World Child Growth Standard mean" [2].

Stunting is associated with negative long term health consequences including physical limitations, retarded cognitive development, increased susceptibility to diseases, increased risk of obesity, and premature mortality [3,4]. The United Nations Children's Emergency Fund (UNICEF) has cited disease and nutrition as the two primary contributing factors to the 23.8% of children stunted globally [5,6]. Previous research has ranked Guatemala fifth worst in the world for child stunting rates, at 49% of all children under the age of five stunted [5]. The purpose of this work is to rank order causal factors to child stunting in Guatemala.

Causal factors of child stunting are diverse, dynamic, and interrelated which deem the issue of stunting a "wicked" problem [7]. To help address wicked problems, systems approaches can provide tools in which to capture the complex characteristics of the system. Primary factors that have been associated with child stunting -- and are present in Guatemala -- include impaired water quality, lack of proper sanitation, insufficient hygiene practices [8], toxins in foodstuffs [9], prenatal health [10,11], caloric and energy intake [12,13], and protein and micronutrient intake [14,15]. Each factor may impact the physical development of a child in a variety of ways and may include 1) limiting the macro- and micro- nutrients that reach the gut, 2) limiting the absorption of those nutrients by the gut, or 3) limiting the immune function that protects a child's gut from infections, among others [16,17]. Based on informal interviews with local non-governmental organizations (NGOs), academics, and government officials working in Guatemala on child stunting, two primary factors currently of interest are fungal toxins in foodstuffs (mycotoxins) and low-levels of chronic exposure to enteric pathogens. Aflatoxin B (AFB), a type of mycotoxin, is produced by the fungus Aspergillus flavus and A. parasiticus and has been classified as a group 1carcinogen by the WHO [18] as it is associated with liver cancer. Previous research has reported on the potential association of AFB and Fumonisin B (FB) with reduced enteric immune function and child stunting [9]. Similarly, enteric pathogens from poor water, sanitation and hygiene (WaSH) practices have been associated with diarrheal occurrences [19], but recently an increase in research has occurred focused on the impacts of enteric infections on environmental enteric dysfunction (chronic inflammation in the gut) [20,21]. To investigate these associations to child stunting within the larger system present in Guatemala, several system analysis approaches were applied to several sets of data from the western highlands of Guatemala.

2. OBJECTIVES

The primary goal of this doctoral research is to advance the fundamental knowledge within the following three hypotheses. Among children in the western highlands of Guatemala between 0 and 5 years of age;

Hypothesis #1: there is a statistically significant association between the severity of the children's environmental enteric dysfunction (EED) and the ratio of the children's height-for-age.

Hypothesis #2: there is a statistically significant association between the children's aflatoxin B exposure level and the ratio of the children's height-for-age.

Hypothesis #3: there is a statistically significant association between the children's aflatoxin B exposure level and the severity of the children's EED.

The secondary goal of this research is to rank order the primary contributing factors to child stunting within the western highlands of Guatemala and a specific set of communities chosen for this studied in Guatemala. These outcomes will allow for improved selection of interventions, both technological and policy oriented, for development professionals including engineers. To complete the identified goals, four objectives were established and a timeline proposed (Table 1):

1. Develop a methodology that improves accuracy of current models representing the causal factors to child stunting

   Improve how data and information can be used with network analysis algorithms (NAA), structural equation models (SEM), and system dynamics models (SDM)

2. Use previously collected data to train and test NAAs on child stunting causations

   Categorize immediate and secondary causal factors; assess their effects on child stunting at a household level and community level; reduce number of potential causal factors to assess; rank-order critical parameters in the system

3. Conduct field assessment using validated survey to test hypothesized correlations between causal factors and child stunting (child height-for-age)

Data collection using a survey and field tests will be conducted at two time-points by the local health center, in San Vicente, Guatemala, and include at least 300 children under the age of five and their mothers

4. Develop geospatial models for water, sanitation, and hygiene based infrastructure barriers to infectious disease transmission. Test models utilizing secondary data from the western highlands of Guatemala

Table 2.1. Anticipated Project Timeline

Obj.	2016				2017			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	Method Devel.							
2		Model Previous Data						
3		Field Analysis						
4					Model & Eval.			

3. LITERATURE REVIEW

Beginning in the 1970s, child stunting, has been promoted as the best long term, national level health indictor [22]. A child is most vulnerable to stunting within the first five years of life [23,24]. Several studies suggest that once a stunted child reaches three to five years of age, the effects are irreversible [25]. The physical growth of a child is complex, but has been found to be associated with the short and long term health as well as interactions that occur in a child's small intestine [26,27]. These associations can be grouped into three categories; first, access to sufficient nutrients; which is highly dependent on feeding practices for the child and community access to nutritious foods [28]. Second, the immune system function of a child; this includes functions such as nutrient allocation to the immune system to fight enteric pathogens, pathogens reducing general absorptive capacity of the intestinal wall (villi), and the passing of nutrients due to chemical imbalances [16,17]. Third, access to the correct nutrients; nutrient needs fluctuate depending on which type of development stage the child is in, while insufficient intake of a particular nutrient can negatively affect child development [29,30]. Many of these enteric problems are hypothesized to be caused by external environmental factors. This provides an opportunity for engineers to engage in identifying the harmful pathways affecting children and to develop barriers to reduce enteric problems in children and therefore reduce stunting. This dissertation presents the development and testing of models to aid in identifying the external pathways affecting child growth and focuses on enteric pathogen transmission, mycotoxin exposure routes, and the subsequent impact on child growth rates in Guatemala.

3.1. ENTERIC PATHOGEN TRANSMISSION

Enteric pathogens can negatively impact a small intestine that is still in development by reducing immune response function and hindering proper development of the microbiome. One of the most widely recognized symptoms associated with increased enteric pathogen loads are diarrheal occurrences [31]. Dehydration due to diarrhea is currently the second leading cause of death for post neonatal children under the age of five [32]. It is also correlated with child stunting; for example in one pooled, nine country analysis, 25% of all stunting was attributed to more than five doubts of diarrhea during the first two years of life [19]. While this correlation is well established in the literature, less is known about the relationship between enteric pathogens and the nutrient absorptive capacity of the intestinal wall [17]. During pathogenic infections, two primary responses occur, 1) T-cells, macrophages and other cells attempt to fight the infection while 2) the villi which line the gut to absorb passing nutrients, recoil. Recent research has reported that with chronic exposure to low levels of pathogens; these nutrient absorbing villi remain recoiled, or blunted, indefinitely [16,20]. Therefore, when children are consistently exposed to unsanitary conditions in and around the home, their ability to breakdown and utilize consumed nutrients can potentially be reduced.

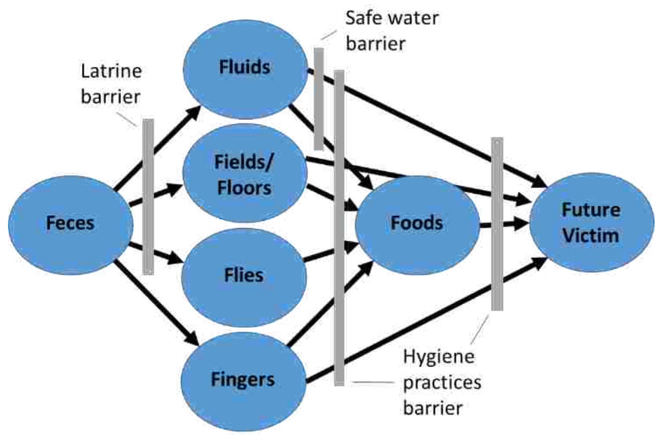


Figure 3.1. The 5F diagram showing the common diarrheal disease transmission pathways.

Numerous studies show that when both children and adults are removed from unsanitary conditions, immune function, intestinal absorption, and growth rates return to normal [33]. Unsanitary conditions refer to an environment where the probability of infectious disease transmission is high, usually due to numerous enteric pathogen transmission routes having increased loads of pathogens. These transmission routes are depicted in a figure called the '5F Diagram' (see Figure 1)[34,35]. This refers to the categories of transmission which include fingers, fluids, foods, floors, and flies. Studies conducted specifically within Guatemala have identified sources of pathogens including water sources [36–39], sanitation facilities [36,37,40], and hygiene practices [36,41,42]. While there are many types of pathogens, sources, and pathways, several general trends have been identified in Guatemala. First is the complex dynamic between the highlands, lowlands and rainy seasons. The lowlands host a less extreme wet and dry season, having rain most of the year. This provides the opportunity for two harvest seasons for farmers, but as a consequence, creates abundant standing water commonly home to water borne pathogens [43]. Second are the cultural habits of rural households that impact the health of a child. Mothers will often carry their child in a sling for the first year of life which reduces exposure to pathogens on the ground, but limits mobility. It is also common practice to begin complimentary feeding before the child is six months of age which increases the number of transmission pathways for enteric pathogens to affect the child [44,45]. Finally, percentages of households who have access to improved water and sanitation facilities are 92% and 78%, respectively [46]. Based on the data and observations from local health workers, there is a high probability that a majority of children are consistently exposed to enteric pathogens through several different pathogen transmission pathways.

3.2. MYCOTOXIN EXPOSURE

There are two types of mycotoxin that have been hypothesized to impact child growth; Aflatoxin B (carcinogen, AFB) and Fumonisin B (FB). Both of these mycotoxins have been found in high concentrations throughout Guatemala [47]. It has been known since the 1970s that high levels of AFB exposure can lead to aflatoxicosis as well as liver cancer [48]. However, within the last 15 years, numerous studies have reported correlations between aflatoxin exposure and child stunting. Wild et al. highlighted six studies conducted since 2002 that found a link between these two variables [9] however, all six were conducted in African countries. Torres et al. have conducted several studies measuring levels of AFB and FB in all departments of Guatemala, finding a range of 0-2600 parts per billion (ppb) with a mean of 63 ppb [47]. The FDA limit for AFB in the United States is 20 parts per billion (ppb) [49]. FB has been shown to affect the development of the neural tube in utero [50], but can also affect child growth [51].

Based on animal models and the few human studies of AFB and FB on child stunting, a set of mechanistic pathways have been hypothesized for the relationship between AFB and FB exposure and the intestinal health of children. The mechanistic theory for AFB and FB is primarily mediated through enteric immune system dysfunction via overstimulation. Currently the two primary hypothesized causal pathways of mycotoxins on child stunting include 1) reduced nutrient absorptive capacity and 2) the modulation of the insulin-like growth factor which has been shown to be associated with child stunting [52,53].

Outside of the body, there exists a multi-level complex system as well. Mycotoxins are given off by fungi that are able to grow in the field (AFB), in storage (AFB & FB), in transport (AFB & FB), and in the market (AFB & FB) [9]. Specifically, within Guatemala, environments differ between highland and lowland communities. This creates a situation where, due to certain market forces, mycotoxin laden maize is grown in the lowlands, but shipped to the highlands, causing multiple exposure routes [54,55]. On the national scale, mycotoxin exposure control is challenging due to the vast weakly regulated transportation system and the lack of source labeling regulations in Guatemala. Poor infrastructure creates longer storage and transport times, while basic pickup trucks used in maize transport are not designed for crop transportation. However, this project will attempt to bridge the gap between the national level and the enteric functions level, by focusing on the household level. Numerous exposure pathways are potentially present within this system and begin either through subsistence maize farming or maize acquired from a market. Subsistence crops can become infected due to misuse of fertilizers and herbicides, cultural harvest/post-harvest practices, weather conditions, poor storage facilities, economic pressures from local maize buyers, and through inhalation if mothers and children work in contaminated work areas [48,56,57]. Exposure of mycotoxins in maize acquired from a market in Guatemala may be due to the food transport duration, original location, the purchase habits of the mother, or the economic status of the household. This project will address two major unknowns associated with mycotoxins; first, the association of mycotoxin exposure to enteric pathogens and child stunting. Secondly, it will test potential exposure pathways within two systems, subsistence farming and market purchases.

3.3. ADDITIONAL SYSTEM FACTORS

Several other critical factors that have been reported as significantly correlated with child stunting in Guatemala include prenatal health [10,58], caloric and energy intake [12,13], and protein and micronutrient intake [59]. Each of these factors are related with breastfeeding and complimentary feeding practices. In total, nutrition plays a significant role in the development of a child however, because of the complex interaction with enteric infections, the understanding of the system may be limited [60,23,61].

3.3.1 Prenatal Health. Prenatal health can be divided into two sections; pregnancy health and multigenerational health. During pregnancy many factors contribute to the development of the child in utero and these factors also vary in importance during each trimester [62,63]. Consumption of foods by the mother has direct impacts on the child and includes proper nutrients, sufficient calories, exposure to enteric pathogens [64], and mycotoxins [65], among others. Critical priorities highlighted by the WHO for pregnancy health include having at least four health center checkups, eating healthier foods, taking iron tablets and other supplements recommended by a health center, and avoiding exposure to insects, among others [66]. Several studies investigated healthy weight gain based on trimester, however, results varied on identifying a priority trimester [67–69]. Finally, access to health facilities and proper delivery facilities reduce mortality and improve the health of both the mother and the child [70]. Multigenerational health factors include physically underdeveloped mother births, underdeveloped birthing organs, and potential epigenetic impacts [71]. Within Guatemala several factors take priority including nutrient consumption, health clinic access, and underdeveloped birthing canal [72–74]. Two emerging topics include mycotoxin exposure [65] and epigenetic effects [75]. Children born in Guatemala are on average halfway stunted at birth [11,76]. There are many factors associated with prenatal health and some play a prominent role in the child stunting problem in Guatemala.

3.3.2 Caloric Intake. According to the Food and Agriculture Organization (FAO) caloric and energy intake refers to the consumption of macronutrients to attain a sufficient level of calories for one day based upon the passage of food into the mouth. There are two processes that regulate what the child actually acquires in the blood stream; the rate of food into the mouth and the rate of utilization within the gut. These processes become even more important during rapid growth periods of children. Breastmilk and complimentary feeding by the mother dictate the rate of food into the mouth of the child, while enteric infections dictate the rate of utilization within the gut of the child. For children, enteric infections can lead to reduced caloric absorption by the intestine as well as a reduced willingness to eat [16,52]. If a child does not retain a sufficient number of calories in the blood stream, fat stores are then drained to provide sufficient energy for growth and fighting infection. This is often related to acute malnutrition and measured through wasting or low weight for height [77]. In Guatemala, the percent of children that obtain the minimum healthy diversified diet – four food groups – is only 36% for children 6-8 months, 49% for children 9-11 months, and 37% for children 12-23 months. The typical diet for families living under the domestic poverty line consists primarily of tortillas and other maize products, potatoes, black beans, sugar, tomatoes, onions, eggs, and coffee [78]. The FAO has identified two primary foci at the community level related to caloric intake; access and stability. These refer to the basic needs of a community in terms of a stable food supply and the ability to purchase these foods [79].

3.3.3 Protein and Micronutrient Intake. Protein and micronutrient intake has become a strong focus for child development. The term 'hidden hunger' refers to people who obtain sufficient calories, but lack particular micronutrients. This is the most common type of malnutrition in Guatemala [78]. Both protein and micronutrients play a significant role in the growth of children and the proper function of their immune system [80,81]. Depending on the type of nutrient (Vitamin A, Iron, Zinc, etc.), the small intestine absorbs them at different locations along its wall into the blood stream. These nutrients are then put to use in one of two general ways within children; either for growth and development or to strengthen the immune system during an infection [77,82]. A child receives a number of significant benefits from breastfeeding including a specific set of nutrients for infants, specific saccharides that initiate particular bacterial growth in the healthy formation of the gut microbiome, and protection from infectious disease through transmission pathway blocking and supplementation of Immunoglobulin A (IgA) for gut health [83]. A significant concern is when the mother does not consume the proper nutrients or sufficient nutrients and is not able to either provide sufficient breastmilk or her breastmilk lacks all the necessary nutrients. The second part of consumption for a child is the complimentary feeding transition. The WHO strongly recommends exclusive breastfeeding until six months of age and then beginning complimentary feeding until the child is two years of age [84]. Proper micronutrients and protein are equally as critical in the complimentary food, but is often what is limited either due to local resources [78] or lack of understanding by the mother [85]. Within Guatemala 49.6% of children are exclusively breastfed until six months, by age two only 46.2% of children are still breastfeeding, and children commonly lack iron, Vitamin A, and iodine [86].

3.4. OTHER POTENTIAL FACTORS

Finally, a number of other factors that are of interest to the public health field and potentially linked to child stunting include epigenetics, ethnical and cultural practices, geographical and logistical systems, and physical and psychological abuse [87–89]. Utilizing the new WHO Child Growth Chart Standards potentially reduces the confounding effect of epigenetics and several studies argue environmental factors within the current generation capture the largest variance of height-for-age changes among children [90,91]. Ethnicity, cultural practices, and logistics will be incorporated into the study design to control for potential confounding effects.

3.5. SYSTEMS ANALYSIS

Systems analysis approaches often have one or more of the following characteristics in common including 1) nonlinearity, 2) feedback loops, 3) time delay effects, and 4) model development [92]. The proper design and application of a systems analysis tool is critical for reliable inferences of the problem being addressed. Structural equation modeling (SEM) is a multivariate statistical tool that allows for a potentially more accurate mathematical design of the real complex system. To improve and compare the validity and accuracy of the set of equations designed to assess the influence of causal factors on child stunting, network analysis algorithms (NAA) and geospatial models will also be utilized.

3.5.1 Structural Equation Modeling. SEM is a statistical technique that has two defining characteristics which provide unique insight into specific systems, factor analysis and path analysis. Latent or hidden variables are concepts that cannot be captured in one observable variable. Common examples of latent variables include the intelligence quotient (IQ), happiness, and even wealth. Outcome (manifested) variables are collected which are hypothesized to be independently influenced by the underlying latent construct or variable. One can think of these observed variables as symptoms a doctor would use to diagnose or measure the severity of an internal illness in a patient [93].

These latent variables can then be integrated into a model which combines observable variables (covariates) and subsequently how both of these types of variables affect one endogenous variable. The second defining characteristic is path analysis, where both direct correlations between an exogenous variable and the endogenous variable can be captured, but indirect correlations through mediating variables can also be captured. This is done by comparing the covariance matrix of a hypothesized set of variables (the model) to the covariance matrix of the data collected to test the model. Fit of the model is measured using four metrics including the chi-square value, the root mean error of approximation, the confirmatory factor index, and the tucker-lewis index. The combination of these two techniques can allow for certain complex systems to be more accurately modeled [94]. Further discussion of specific methodologies are presented within each Paper.

The aim of this study is to utilize this modeling technique to improve the accuracy and applicability of our current models used in diagnosing problems in the child health sector of Guatemala. Data from two time-points will be used to test 1) cross sectional SEMs and 2) two time-point SEMs. The statistical methods used to develop and test each of these types of SEM applications are the same. This application of SEM is used in Papers II, III, and IV to build height-for-age (stunting) models, EED models, and AFB models.

3.5.2 Network Analysis Algorithms. NAA has become more popular with the advancement of computational power in computers and the increased access to large amounts of data. There are a number of different types of algorithms used in NAA but the weighted correlation algorithm will be utilized in this dissertation. Secondary data sources available from Guatemala include regional household surveys from the US Agency for International Development and Guatemalan Government from 2012 and 2013. Utilizing these data, a directed path algorithm within the weighted correlation algorithm family will be applied to force a large number of variables to decide how to hierarchically associate themselves with child height-for-age in the most optimal way possible. This is theoretically different from SEM, as the hypothesized SEM model represents a notable-null hypothesis approach ($h0 \neq 0$) however, the NAA can handle a larger number of variables and data points. This output from the NAA, field observations, expert opinion, and a literature review will be utilized to inform the hypothesized SEMs.

3.5.3 Geospatial Modeling. Finally, geospatial modeling will be conducted in collaboration with the US Agency for International Development (USAID) and the Guatemalan Government. Geographic data from 2012 will be utilized from the western highlands of Guatemala to build regional SEMs for five departments including Huehuetenango, San Marcos, Quiche, Totonicapán, and Quetzaltenango. The geospatial models will focus on infectious diseases transmission barriers and potential negative outcomes such as diarrhea, EED, and child stunting. Once built, geographic data from 2013 will be utilized to test and potentially validate all SEMs and investigate 1) regional similarities in the western highlands, 2) regional trends between groups of departments, 3) and site specific characteristics for each department. If other data is available from the community site this study utilizes or other studies sites, the applicable department model will be tested against that data. This methodology for the development and validation of these models is presented in Paper V.

4. PRELIMINARY FINDINGS

4.1. SEM AND ENTERIC INFECTIONS IN GUATEMALA

Recent research by our team demonstrated the applicability of the primary modeling technique, structural equation modeling, on causal factors to diarrheal occurrences among children from Guatemala and Brazil. Both studies utilized the SEM methodology to rank order variables associated with diarrheal occurrences. These studies provide the foundation for the EED model. The SEM methodology will also be used to analyze the mycotoxin exposure system and the child stunting system.

Divelbiss et al. conducted an evaluation of the effectiveness of a biosand filter to reduce diarrheal occurrences in households located in the Ixcan region of Quiche, Guatemala [99]. The team hypothesized an initial model based on field observations and literature reviews. Three rounds of data collection were conducted to test and improve the model. Once fit statistics showed adequate fit of the data to the model, parameter estimates could then be evaluated. Figure 2 shows the final model and associated parameter estimates for each relationship within the model.

The model depicts the significant relationships with diarrheal occurrences and the significant relationships with a household's ability to operate and maintain their filter properly. While the filter did help reduce diarrheal occurrences (-0.119), household education (-0.170) and improved water source (-0.169) were most important. For operating and maintaining a filter, only soap present in home correlated positively, suggesting there are associations with hygiene practices and filter operation practices. Additional water treatment had the largest negative effect on operating the filter properly. This method was then validated in subsequent work utilizing the same tool, but adjusted for a different environment in Brazil.

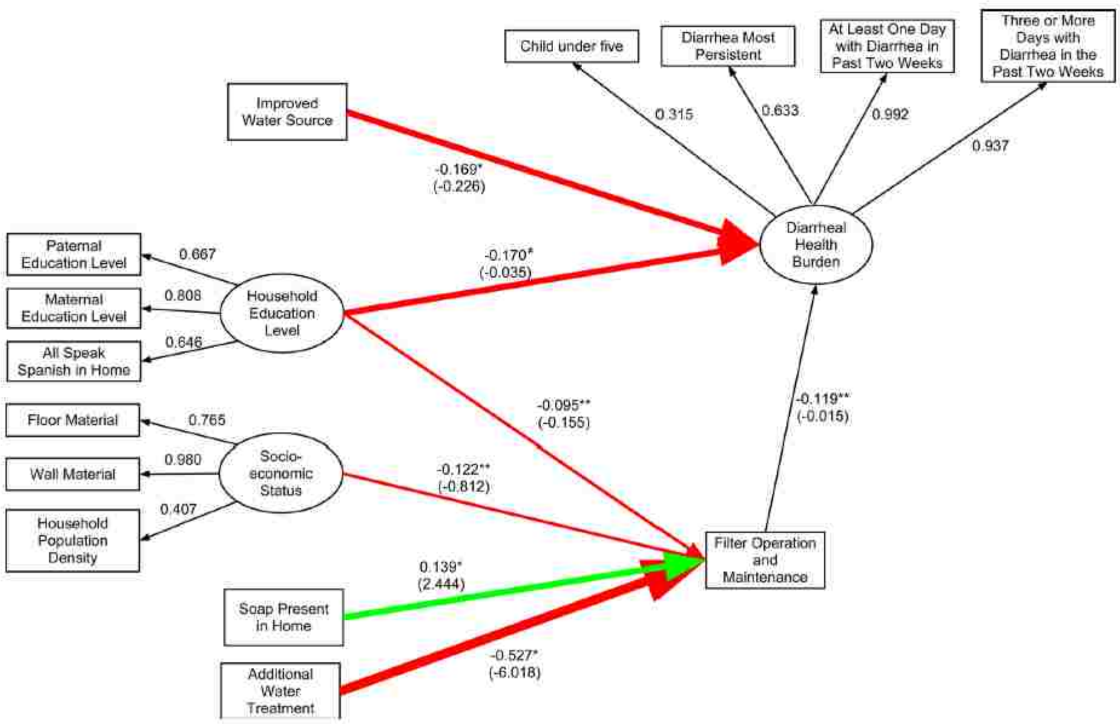


Figure 4.1. Final Guatemala hybrid model (structural and measurement). The standardized and unstandardized (listed in parentheses) parameter estimates are listed next to the associated pathway. * $p < 0.05$, ** $p < 0.1$, # $p < 0.15$, measurement error terms (e) were removed to reduce congestion. Ovals are latent variables, rectangles are observed variables, and arrows depict hypothesized relationships. Weight added to arrows for emphasis; color indicates direction of influence, red is negative influence, green is positive influence.

4.2. SEM APPLICATION IN BRAZIL

For the study conducted in Brazil the SEM model and associated survey from Guatemala was contextualized for the state of Para [100]. Three villages along the Amazon River northwest of Santarem, Para, Brazil were studied. Two iterations of data collection were needed to reach a parsimonious model. Figure 3 depicts the relationships within the system impacting both filter operation and maintenance and diarrheal occurrences [101]. The results showed that the filter had little impact on diarrheal occurrences, while household education and sanitation facilities had the largest beneficial effect sizes. One possible reason for the low impact of the operation and maintenance of filters on diarrheal occurrences may have been due to the strong negative impact from additional treatment technologies. Previous research has reported that too many treatment technologies may overwhelm the user, reducing overall disease protection.

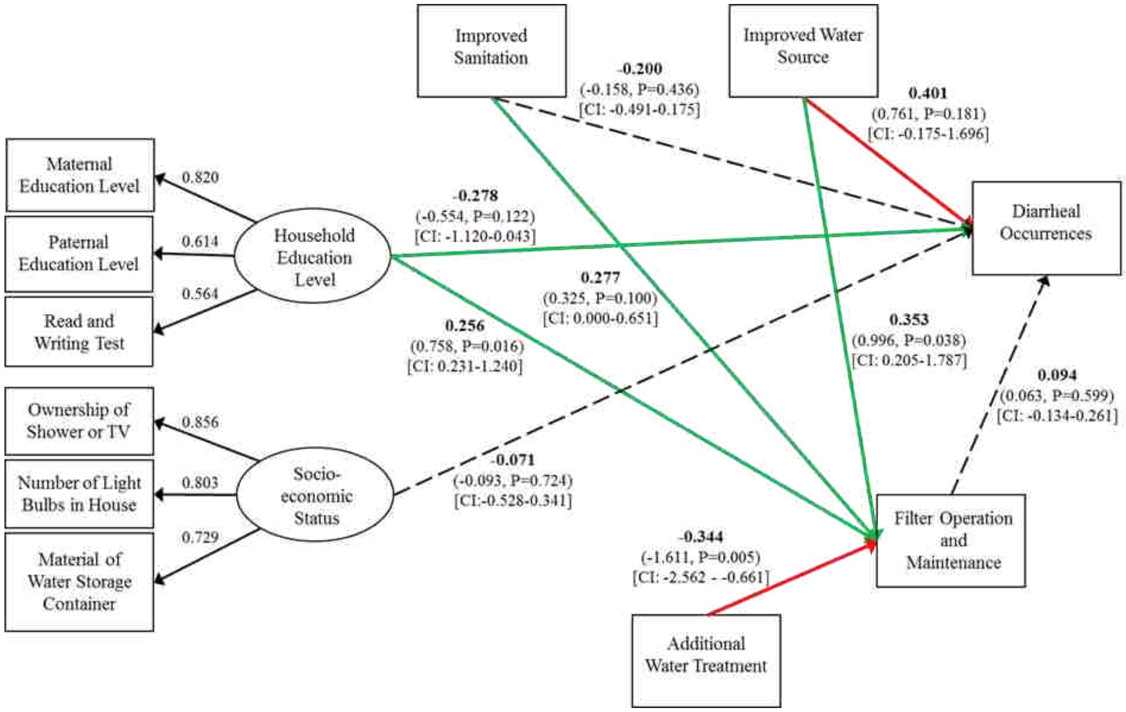


Figure 4.2. Final Brazil hybrid model (structural and measurement) with final parameter estimates of hypotheses. Dashed arrows identify insignificant relationships (p>0.20). Standardized estimates given in bold, unstandardized estimates are in parentheses with p-values and confidence intervals. Overall model fit was good (χ^2 p>.617; RMSEA = .000 [CI: 0.000-0.093]; CFI = 1.00; TLI = 1.08).

In addition to the SEM data and analysis, secondary data and several basic statistical techniques were applied to confirm findings within the Brazilian and Guatemalan studies [102]. Mahalanobis-Taguchi Strategy (MTS), Canonical Correlation Analysis (CCorA), and Latent Factor Regression (LFR) were used to analyze data collected by the Demographic and Health Survey program in Brazil and Guatemala.

The secondary analysis confirmed several key relationships identified in the SEMs, but also identified several other variables, not included in the SEM that should be considered in future work to better explain the variance in diarrheal occurrences. For Quiche, Guatemala, factors identified as significant included education level of parents (MTS, CCorA, LFR), ethnicity (CCorA, LFR), sex of household head (CCorA, LFR), and water source (MTS). For Para, Brazil, factors included education level of parents (CCorA, LFR), sanitation (CCorA, LFR), socio-economic status (MTS, CCorA, LFR), and household social structures (MTS, CCorA, LFR).

These studies demonstrate the applicability of a set of tools in the assessment of enteric infections. The use of SEM was demonstrated in the application to assessing the efficacy of biosand filters in Guatemala (Divelbiss et al. 2013) along with its applicability in different environments (Voth-Gaeddert et al. 2015a). The team also demonstrated the utilization of multiple statistical techniques (MTS, CCorA, LFR) in analyzing enteric infections (Voth-Gaeddert et al. 2015b). In this dissertation, the aim is to investigate an enteric infection (EED) SEM, along with two AFB SEMs. Finally, a child height-for-age (stunting) SEM will be tested to investigate the hypothesized effects of EED, aflatoxins, and nutrition. Furthermore, this dissertation aims to expand on the use of integrating statistical techniques and introduce geospatial SEMs as another technique to improve the understanding of the complexity of child stunting in Guatemala.

5. OUTLINE

The results are reported in the format required by the specific journal in which each manuscript was originally submitted. This means that each Paper includes an introduction, methods, results, and discussion section at a minimum specifically written for that journal. In this dissertation, Papers I – V are five manuscripts while the second section is a brief discussion and conclusion of the full dissertation. To provide guidance for the reader the rest of the dissertation is outlined below. As the introduction to the dissertation topic has been provided above, a cohesive methods and expected results section is provided below.

As briefly highlighted in the objectives for the dissertation a three-step methodology was utilized to test the hypotheses. This included 1) the application of network analysis algorithms to larger data sets, 2) the development and testing of SEMs with field data, and 3) the development and testing of SEMs from regional data.

First, data from USAID's Food for Peace Title II Baseline Survey was acquired, aggregated, and prepared for analysis. 2,103 children were included in the data set as well as 87 variables which had been selected based on the WHO recommendations for causal factors to child stunting. A weight correlation network analysis algorithm was applied to the data and several spanning tree diagrams were produced based on the strength of relationship between child height-for-age z-score and the other 86 variables. Outputs included a tree diagram for the child height-for-age z-score, a tree diagram for child diarrheal prevalence, tree diagrams for child height-for-age z-score for different age categories, and tree diagrams for child height-for-age z-score for different levels of stunting severities. Results are presented in Paper I and the first half of Paper II.

Second, information was aggregated from the network analysis output, field observations, a literature review, and informal interviews with locals and experts to hypothesize a set of SEMs. These included 1) the three-way interaction between AFB exposure, EED symptoms, and child height-for-age, 2) causal factors to low child height-for-age, 3) causal factors to increased AFB symptoms, and 4) causal factors to increased EED symptoms. Data was collected from the community of San Vicente, Totonicapán in two field campaigns, October 2016 and February 2017. This data was then applied to the SEMs in a confirmatory approach to test the hypothesized correlations. Outputs included model fit indices, parameter estimates, and p-values for the four SEMs listed above. Paper II presents the approach to hypothesizing, testing, and results of the child height-for-age and the EED SEMs. Paper III presents the results of the three-way interaction SEM and Paper IV presents the results from the causal factors to increased AFB symptoms SEMs.

Third, in order to test the scalability of the SEM approach, regional data was utilized to develop and test five department (or state) specific EED SEMs. Additionally, three different health outcomes (diarrhea, EED, and height-for-age) were incorporated for a total of 15 SEMs (three per department). The USAID Food for Peace Title II Baseline Survey 2012 was utilized to develop the regional SEMs (exploratory approach) and the USAID Western Highlands Integrative Program Baseline Survey 2013 was used to test the SEMs (confirmatory approach). Finally, as all input variables in the models were WaSH infrastructure related, the identified set of transmission pathways related to the individual WaSH infrastructure variables (as identified by a literature review) were used to provide suggestions on specific transmission pathways of importance for that department and health outcome. Outputs included statistically significant WaSH variables for each department and health outcome as well as potentially important transmission pathways for each department and health outcome. These results are presented in Paper V.

PAPER

I. IMPROVING HEALTH INFORMATION SYSTEMS IN GUATEMALA USING WEIGHTED CORRELATION NETWORK ANALYSIS: DEVELOPMENT AND APPLICATION OF NETWORK ALGORITHMS FOR UNDERSTANDING CHILD STUNTING

Published in IEEE Global Humanitarian Technologies Conference 2016 Proceedings

Lee Voth-Gaeddert1, Devin Cornell2

1 Civil, Architectural, and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO, USA, lv6w3@mst.edu

2 Department of Sociology, University of California, Santa Barbara, Santa Barbara, CA, USA, dcornell@umail.ucsb.edu

ABSTRACT

Guatemala has the fifth worst child stunting prevalence – low-height-for-age – in the world, at 49%. Child stunting is associated with negative short and long-term health effects and the contributing factors are complex, interrelated, and potentially non-linear. Current health information systems (HIS) in Guatemala are disaggregated, overly complex, and have limited scalability. This paper demonstrates the use of weighted correlation network analysis to visualize and explore data in a way that provides useful information for future HIS. The methods generate a holistic causal factor model for stunting that explores how cofactors relate to stunting and each other. The demonstration here is based on a Guatemala regional data set obtained from the USAID Open Data Website. First, the overall correlation network structure is observed and compared to generalized structural models proposed by the WHO and USAID. Next, quantile comparisons are performed using the outcome variable z-score height-for-age, and distinct child age groups. The comparisons demonstrate how these networks can be used as an extension of widely used methods while also providing advantages that are important for exploratory analysis. This work is an important first step in evaluation of a novel analysis method for health information systems currently being developed in Guatemala.

1. INTRODUCTION

In 2015, the United Nations members ratified 17 Sustainable Development Goals (SDGs) set to be achieved by the year 2030. SDG 2.2 states that, “by 2030, end all forms of malnutrition, including achieving, by 2025, the internationally agreed targets on stunting and wasting in children under 5 years of age. [1]” Stunting is defined by the World Health Organization (WHO) to be a characteristic of a child that is two standard deviations (SD) below the mean height for his or her age [2]. Guatemala currently is fifth worst in the world in terms of stunting of children under the age of five at 49% [3]. Many short and long-term consequences have been identified and include increased susceptibility to diseases, stagnant cognitive development, reduced physical stature, increased risk of obesity, and premature mortality [4], [5].

The causal factors that have been identified for stunting are broad, interrelated, dynamic and potentially non-linear [6]. The WHO has provided guidance towards graphically describing the multi-layered system in their publication, “Childhood Stunting: Context, Causes and Consequences” [7]. Guatemala hosts a diverse environment, geographically, ethnically, politically, and climatically which creates challenges to provide health services to all citizens. General access to resources for the population is low and includes limited medical personnel and equipment for health centers [8], [9]. With so few resources available, the importance of useful health information for targeting resources at a community level is critical [10].

Health information systems (HIS) are a key element in providing complete health systems to overcome the complicated challenges developing countries face. The four elements of a HIS are defined by the WHO to be data generation, compilation, analysis and synthesis, and communication and use [11]. This study will focus on the improvement of ’analysis and syntheses’ as well as ’communication and use’. These systems are used to collect and analyze data to support decision-making on health interventions. The analysis methods currently used in the field often reflect the questions that decision-makers had prior to the data being collected [11]. Furthermore, it has been reported that HIS in Guatemala are often fragmented across organizations in both analysis and synthesis and the method of communicating and utilizing results. This sometimes leads to the loss of a holistic picture of the problem [8]. A common interface with more exploratory capabilities is needed to standardize communication while retaining all information that may be useful for the problem.

This work attempts to meet the analysis and communication needs of the Guatemala HIS by presenting novel network based methods that also use tools for visualization and exploration that current systems lack. Although these methods are novel for HIS, they take inspiration from increasingly popular methods in gene cofactor expression as well as new tools for visualization and exploration of networks [12], [13]. The methodology used involves several steps: 1) creation of a correlation network where nodes are measured indicators from the survey and edges are correlations between them, 2) transformation of those edge weights for desired analysis, and 3) creation of a shortest-path spanning tree centered on the outcome indicator 'z-score height-for-age' (abbreviated zhaz). The resulting spanning tree can be output as a ".gexf" file which can be opened in a graph visualization program such as Gephi. Then, comparisons can be made across quantiles of the outcome variable or cofactors by examining structures of the resulting trees.

The tools presented here allow the user to look at how all cofactors are related to outcome variables in a holistic way. The raw correlation network by itself is too much information for a person to consume without significant effort, but the outcome-centered spanning tree allows for easy observation of strong causal pathways through all possible cofactors. In addition, the ability to visualize these pathways and interpret structural differences could change the way we think about causal analysis. The tool and methods here are still in early stages of development but they appear to address both the analysis and communication problems currently faced by the HIS in Guatemala.

2. METHODS

Surveyed households were primarily agrarian farms selected for the "Baseline Study of the Title II Development Food Assistance Programs in Guatemala" [14]. Multiple tables of the dataset were combined to make single a table where each row was a child and each column was one of 87 indicators relating to Child Health, Household Description, Maternal Health, Sanitation, Breastfeeding Information, and Agricultural Practices [14]. [see supplementary material for variable descriptions]. As the height-to-age measure was the outcome variable and breastfeeding has been identified as a critical factor to height-to-age [4], children with missing data in either of these two variables were eliminated for this analysis (n=2103 remaining). The goal of these methods is to take the table representing encoded survey data and convert it to an interactive visualization that can help aid workers understand relationships between cofactors and z-score height-for-age.

Data analysis algorithms were built using the Python programming language with the Numpy, Pandas, and NetworkX libraries [15]–[18]. This choice of programming language and tools was made so that a future web application could be built without a large change in the code. The python algorithms take the survey data table and output a ".gexf" graph file which can be opened in Gephi [12]. Gephi is used as a graph visualization program taking raw graph data (with node and edge attributes), and using it to color and position nodes and edges in a 2D space.

First, the encoded survey data file (in ".xlsx" format) is read in and converted to the Pandas DataFrame format for manipulation. Next, a complete undirected graph is constructed where each node corresponds to a specific question in the survey, which we will assume is a random variable. Several types of variables were not added as graph nodes: nominal variables (sex, location, survey date, etc.), derived indicators (household diversity score, total consumption, poverty score), and outcome variables (body mass index, weight-for-age, and weight-for-height, and weight-for-age). Although these were not used as nodes in the graph, they will be used later for comparison. In this case, the outcome variable will be considered to be z-score height-for-age, the primary indicator used to measure stunting.

The conversion of encoded survey responses into a correlation network allows only correlation information between each response to be retained. Edge weights w^c were added to each undirected edge to represent the correlation coefficient between the connecting variables. This correlation weight between arbitrary variables i, j for $j \neq i$ is given as $w^c_{i,j}$ in Equation 1.

$$w^c_{i,j} = \rho_{i,j} = \mathrm{corr}(v_i, v_j) \quad (1)$$

Pearson's correlation coefficient was used between scalar variables and Spearman's rank correlation coefficient was used between pairs with ordinal variables.

Data that was missing from the survey was simply omitted from the correlation calculation. Missing data was in all cases below 15% of total entries, and the most affected topics were those relating to farming. This could cause a slight bias in the correlations towards other variables which correlate with the missing data entries, but given the small number of missing values this was deemed insignificant.

The use of a correlation graph stems from the need to understand relationships between all observed variables instead of only the direct relationships between cofactors and outcomes. As stunting has been shown to be a very multifaceted problem [19], [20], it is important to consider multiple causal pathways that could be contributing to this issue.

Although the raw correlation graph contains the most obviously useful information about the inter-related variables, further transformation is needed to understand how covariates affect the outcome variable z-score height-for-age while considering the complexity of the situation. An approach is taken to orient the graph into a tree where ’zhaz’ is the root node and all other variables are descendants of that root. To organize the nodes into the tree structure, a transformation of the correlation edges is needed. Let wp be new weight values for the transformed graph shown below.

$$w^{p}_{i,j} = |w^{c}_{i,j}|^{-\beta} = |corr(v_i, v_j)|^{-\beta} \quad (2)$$

The graph with edge weights wp is one where smaller edge weights correspond to larger correlations and the parameter β will accentuate differences between correlations (more on that later). In network literature, this is often referred to as a ’soft thresholding’ [13]. These weights can be considered as the relative ’closeness’ of two variables based on their correlation. A graph with these properties is convenient for observing shortest path and centrality measurements. In this case, the shortest path algorithm will be used to create a spanning tree using only edges that lie on a shortest path between ’zhaz’ and every other variable. The result is a tree topology that describes the relationship of each variable with ’zhaz’ while taking into account other correlated variables.

The motivation for using the shortest path can be observed by analyzing a simple connected undirected graph with four nodes v_1, v_2, v_3, v_4 (as shown in Figure 1) representing four correlated random variables. Assume that although the graph is connected, the correlation $\rho_{3,4} = 0$ and so corresponding weight $w_{3,4}$ = inf and thus it was not drawn in Figure 1. If we designate v4 to be the node associated with the outcome variable, then we are trying to best understand how variables associated with nodes v_1,v_2 and v_3 'affect' that outcome. We can use the same weight expression w^p given above, and use $\rho(i, j) = corr(v_i, v_j)$ to represent the correlation coefficient between variables associated with v_i and v_j .

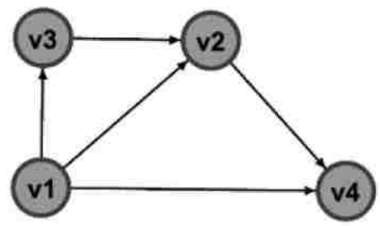


Figure 1. Simple undirected four-node graph.

Weights in this transformed graph are given by $w_{i,j} = |\rho_{i,j}|^{-\beta}$, and so the three possible path distances for variable v_1 are given in Equations 3, 4, and 5. Notation for paths and associated distances will be given through use of p and d with subscripts respectively. A path connecting nodes v_1 and v_3 through v_2 will be given as $p_{1,2,3}$ and the associated distance will be $d_{1,2,3}$. The notation for the shortest path between arbitrary nodes i and j will be $p^{sp}_{i,j}$ and its distance $d^{sp}_{i,j}$. All possible paths connecting v_1 and v_4 can be enumerated for the graph in Figure 1 as $\{p_{1,4}, p_{1,2,4}, p_{1,3,2,4}\}$ and thus $p^{sp}_{1,4}$ must come from this set. The shortest path algorithm is reduced to a selection from one of the alternatives presented in equations 3, 4, and 5.

For $p_{1,4} : d_{1,4} = |\rho_{1,4}|^{-\beta} = w^p_{1,4}$ (3)

For $p_{1,2,4} : d_{1,2,4} = |\rho_{1,2}|^{-\beta}+|\rho_{2,4}|^{-\beta} = w^p_{1,2}+w^p_{2,4}$ (4)

For $p_{1,3,2,4} : d_{1,3,2,4} = |\rho_{1,3}|^{-\beta} + |\rho_{3,2}|^{-\beta} + |\rho_{2,4}|^{-\beta} = w^p_{1,3} + w^p_{3,2} + w^p_{2,4}$ (5)

The shortest path algorithm will calculate the shortest path distance $d^{sp}_{1,4}$ from v_1 to outcome variable v_4, which is shown by Equation 6.

$d^{sp}_{1,4} = \min\{d_{1,4}, d_{1,2,4}, d_{1,3,2,4}\}$ (6)

The algorithm produces path $p^{sp}_{1,4}$ that connects v_1 and v_4 using the smallest possible distance. The distance calculation is obviously a function of all paths in the graph, but is also a function of the soft threshold parameter β [13]. Ignoring β for a moment, observe that our shortest path selection should allow us to understand which possible causal pathway is most significant by removing edges not included in the shortest path. If this is performed between v_4 and each other variable in the original graph with weights w^p, we can leave a spanning tree that includes only edges that appear on one of the shortest paths. The spanning tree represents the smallest possible distance between every node and the outcome variable on the transformed graph.

The effect of parameter β can be examined by looking at the shortest path selection. As a hypothetical assume that weight $w_{1,4}$ is very small compared to the other weights in the graph. If $w_{1,4}$ is the smallest weight and thus $\rho_{1,4}$ is the largest correlation in the graph, then the selection is easy: $d_{1,4}$ will be the shortest path regardless to the other weights and regardless of the parameter β. Now assume an alternative: that $\rho_{1,4}$ is larger than all of the other correlations except for $\rho_{1,2}$ and $\rho_{2,4}$ (discount v_3 for simplicity). The selection of either $p_{1,4}$ or $p_{1,2,4}$ as the shortest path depends on the inequality $w_{1,4} < w_{1,2} + w_{2,4}$ or equivalently $|\rho_{1,4}|^{-\beta} < |\rho_{1,2}|^{-\beta} + |\rho_{2,4}|^{-\beta}$ (truth implies $p^{sp}{}_{1,4} = p_{1,4}$). The assumption $\rho_{2,4} \leq \rho_{1,2} < \rho_{1,4}$ implies that for any arbitrary β, $|\rho_{1,4}|^{-\beta} < |\rho_{1,2}|^{-\beta}$ and $|\rho_{1,4}|^{-\beta} < |\rho_{2,4}|^{-\beta}$.

It is obvious from Equation 2 that a larger β implies a smaller wp (because $\rho < 1$), but it is also true that a smaller ρ will cause the corresponding wp to be affected by β more significantly. By decreasing β, eventually the sum $w_{1,2}+w_{2,4}$ would exceed the value of $w_{1,4}$ and thus $p_{1,2,4}$ will become the new shortest path. This result means that in order for a given path between v_i and v_j to be the shortest path, all of the associated correlations must be shorter than the direct path $p_{i,j}$. As β increases towards infinity, the causal pathway spanning tree actually approaches the minimum spanning tree of the transformed graph. As β decreases towards zero, the causal pathway spanning tree reduces towards a tree of depth 1 where every cofactor is a leaf node whose parent is the outcome variable.

Although these statements require further proof, the proofs are not necessarily needed for the analysis to be useful. In this case, β can simply be thought of as a parameter that determines the degree to which the variables are structured around zhaz. It was experimentally demonstrated in this work that a decrease in some arbitrary β will result in a 'less' structured tree with more leaf nodes and more centrality given to the outcome variable, and an increase in an arbitrary β will result in a 'more' structured tree with fewer leaf nodes and less centrality given to the outcome variable. An arbitrary β may be more or less useful depending on the type of causal understanding and assumed interconnectedness desired. All results shown in this work were computed with $\beta = 2$, and it appeared to show a reasonable balance between structure and centrality of outcome variables that was appropriate for the analysis.

Initially a single stunting-centered spanning tree was created using all of the data, but then analysis was performed using only divided quantiles of specific variables. Quantile separation was performed on three categories of the stunting outcome variable ’zhaz’ and four categories of child age. The data was also split into separate trees for the male female differential and the time of interview [19], [21]. Generating separate trees for different quantiles of these variables will reveal structural differences in the causal factors for stunting as these factors are varied.

In order to compare trees from separate quantiles, the shortest path distance $d^{sp}_{i,j}$ across quantiles was used. A table was generated for each variable on which quantile analysis was performed. This table consists of any variables which were the top 10 most correlated with stunting in any of the quantiles. The variables were then sorted according to the variance of $d^{sp}_{i,j}$ across the quantiles. The end result is a table that prioritizes variables that have a strong connection to stunting but which also vary significantly across quantiles.

The software created and demonstrated in this work is designed to provide novel analysis important for the creation of a country-wide health information system. Future work is needed to implement this system on a live connected system, but the usefulness of this analysis on a real dataset has been demonstrated.

3. RESULTS

The sample population included a total of 2,103 children, of which 1,103 were males and 1,000 were females. 80.5% of them were considered stunted by WHO standards. 60.2% of the mothers of the children reported their child having diarrhea in the past two weeks. Finally, 16.9% of households reported having gone without food for a day within the past month.

3.1. ZHAZ Spanning Tree

The initial zhaz spanning tree output graphically displays the structure of the data utilizing the algorithms discussed above. Figure 2 displays the example generated from USAID's data. As there are 79 potential causal factors modeled, variables identified as less than two nodes (distance from zhaz < 3) from the zhaz score are specifically labeled. Additionally, the location of specific groupings of variables are identified and subsequently discussed. This provides the user with an understanding of how certain sections (sustainable agricultural practices, family demographics, ORTs and Diarrhea, food consumption, etc.) interact in specific situations. Appendix 1 provides a fuller description of the common variables.

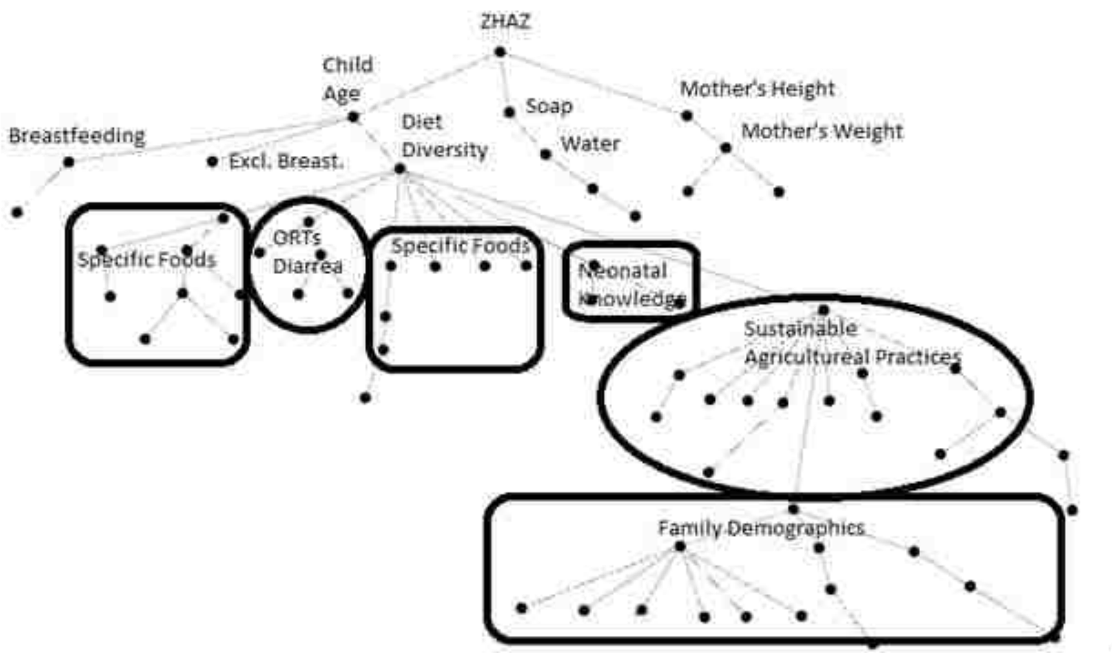


Figure 2. The full spanning tree generated by the shortest path algorithm displays the overall structure of the data acquired from USAIDs online data repository. Nodes closer to the ZHAZ node are considered to have a bigger effect. Variables farther from the ZHAZ variable (less direct impact) are identified under a theme (e.g. sustainable agricultural practices, family demographics, etc.).

For the zhaz spanning tree, variables with a distance of two (i.e. nodes directly connected) from zhaz included mother's height, soap present at hand washing stations, and age of child in months. The variables with a distance of three (i.e. nodes mediated by a second node) included the mother's weight, presence of water at hand washing station, practice(d) exclusive breastfeeding, current breastfeeding status, and diet diversity (DD) score.

The food grouping (i.e. aggregated under diet diversity in this model), including a range of reportedly consumed foods, was closely related to zhaz (consistently two to three nodes away). The food grouping, besides specific foods, also included the use of oral rehydration therapy and reported bouts of diarrhea. Sustainable agricultural practices (SAP) grouping was directly linked to diet diversity. Lastly, the family grouping was the farthest from the zhaz score and connected through the SAP group.

3.2. Quantile Analysis of Stunting

The quantile analysis provides a perspective of the data that utilizes the levels of stunting to generate the model (not stunted: -2 SD+, stunted: -2 to -3 SD, extremely stunted: - 3 SD-). Figures 3, 4, and 5 display all three spanning trees for the different quantiles truncated after the second node for simplicity. The first quantile included children with a zhaz score greater than -2 SD or those children classified as not stunted (see Figure 3). The first level of nodes for this quantile included soil conservation used, the mother's height, issues with maize harvest, potatoes consumed, and age of child. The second level nodes include sustainable agriculture practices, mother understands warning signs of a sick child, the mothers weight, availability of water at nearest source, household language, issues with disease or pest in maize, food deprived in past month, other fruits consumed, meats consumed, vegetables consumed, ORTs, currently breastfeeding, usage of exclusive breastfeeding, diarrhea present in past two weeks, diet diversity score, and water available at hand washing station. The food grouping was split into two groups but was related to the zhaz score (diet diversity). The SAP group was at the second level, while the family group was the farthest from zhaz.

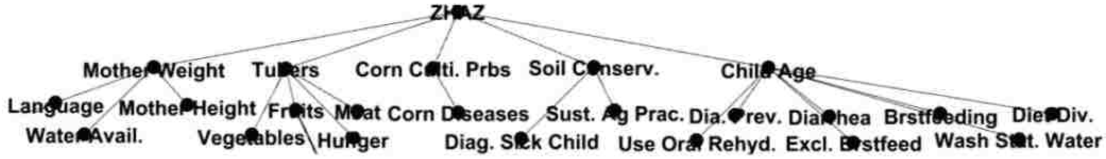


Figure 3. Causal zhaz-centered tree not stunted model which identifies variables most important to child growth rates for children that were not stunted (above -2 SD).

The second quantile of children were those between -2 and -3 SD (see Figure 4). The first level of nodes included the mother's height and age of the child. The second level of nodes include the mothers weight, usage of exclusive breastfeeding, currently breastfeeding, currently pregnant, ORTs, and DD. The food grouping had a large number of nodes and was in the second set of nodes, while the family group was in the third and the SAP group was the farthest removed.

Figure 4. Causal zhaz-centered tree model from children that a were classified as stunted (-2SD to -3SD ZHAZ).

Finally, the worst quantile of child stunting captured any child less than -3 SD (see Figure 5). First level nodes included the mother's height, currently breastfeeding, and age of the child. The second level nodes included the mothers weight, household language, currently pregnant, diet diversity, usage of exclusive breastfeeding, and total number of children in the household. Again, the food grouping played a significant role in the model at the second level, followed by the SAP grouping at the third, and the family grouping at the fifth.

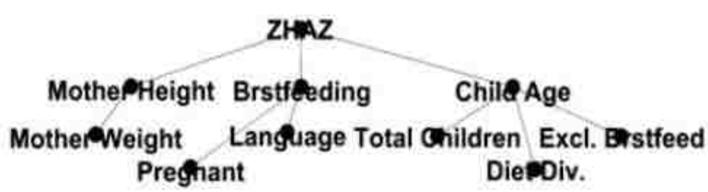


Figure 5. Causal zhaz-centered tree model from children who were classified as extremely stunted (-3SD and below).

3.3. Age-Specific Stunting Factors

Next, the tool separated children based on age to investigate key contributing factors to child stunting within specific age ranges. Figures 6, 7, and 8 display the truncated spanning trees for the three age categories in months; 0-6, 7-12, and 13-17.

The age group 0-6 months had nine first level nodes and fifteen second level nodes (see Figure 6). From the full model the food grouping was divided but was both at the first and second level to zhaz, while sustainable livestock was at the first level. Family and SAP groupings were three and four nodes away, respectively.

Figure 6. Causal zhaz-centered tree model showing the first two levels (for simplicity) of causal variables for children 0-6 months of age.

The age group 7-12 months had six first level nodes including age of the child, improved maize storage, improved animal pens, consumed cheese products, food deprived in past month, and the mother's height (see Figure 7). The second level of nodes had ten variables.

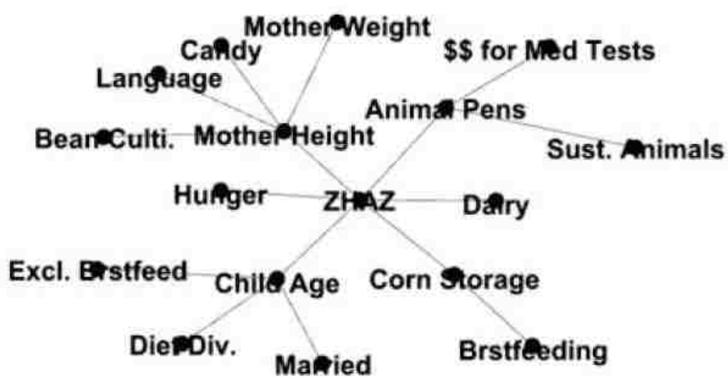


Figure 7. Causal zhaz-centered tree model from children 7-12 months of age.

The age group 13-18 months had seven first level nodes including presence of water at hand washing station, mother's height, spent money on home repairs, total children in household, language, foods made from beans, nuts, lentils, etc., and age of child (see Figure 8). The only grouping that naturally grouped together was the food group, all others were disaggregated and far removed.

Figure 8. Causal zhaz-centered tree model from children 13-18 months of age.

Finally, an analysis was conducted to identify the top five variables that significantly changed over the four quantiles. Table 1 displays these variables along with scores for each quantile. The value represents the importance of the variable to zhaz in a particular quantile (the lower the value the more important the variable). These variables included water for hand washing, soap for hand washing, exclusive breastfeeding, the mothers age, and language, in order of variability (as measured by the standard deviation of the scores across quantiles). For example, as the quantile increases in age, soap for handwashing suddenly becomes very important, specifically in the 13-18 month's age category.

Table 1. Differences in causal structure across ages (in months). The lower the value the more important that variable at the given time. Ranked based on standard deviations.

Variables	0-6	7-12	13-18	19-24	Std. Dev.
Water for Handwashing	5.6	35.2	2.8	9.6	8.9
Soap for Handwashing	8.5	30.9	9.6	3.2	7.6
Exclusive Breastfeeding	4.8	1.9	12.1	6.7	1.7
Mother's Age	2.8	8.5	4.9	9.1	0.8
Language	2.1	1.1	4.2	4.1	0.8

3.4. Child Gender and Seasonal Variations

To investigate potential gender differences, graphs for male and female were generated. Both graphs resembled the structure of the primary zhaz spanning tree graph. There were no changes to the top ten significant variables when gender models were compared. Similar results were obtained when investigating potential differences in data collection times (during the rainy season and during the dry season). Only the mothers age dropped out of the top ten significant variables during the dry season and was replaced by consumption of beans.

4. DISCUSSION

Analyzing large amounts of data creates challenges in reporting and interpreting results. This tool offers a platform in which to begin a more multidisciplinary approach to child stunting, both as a health practitioner and as a researcher. It will only be through an iterative process of model development that will provide the needed set of tools for effective change.

4.1. ZHAZ Spanning Tree

The zhaz spanning tree provided information based on all information across the region for all ages of children. The emergent structure of the data generally follows hypothesized relationships from the literature. The different levels of nodes in the zhaz spanning tree were similarly grouped compared to the major categories of the WHO graphical models. These categories include breastfeeding practices (exclusive breastfeeding and breastfeeding), WaSH practices (soap and water present at hand washing stations and water availability), micronutrient and protein consumption (DD and subsequent variables), caloric intake (DD), and prenatal health (mother's height and weight). While a systems level validation of the hypothesized relationships identified by the WHO is useful, the aim of this tool is to provide a platform for hypothesis development of potential critical relationships and, most importantly, the testing of these hypotheses. This will become the validating step to the systems level analysis. The paper demonstrated this through the quantile stunting analysis and the age specific analysis.

4.2. Quantile Analysis of Stunting

By using a quantile analysis, different data structures were created by the tool for each category of child (not stunted, stunted, and extremely stunted). The data and subsequent graphs showed number of interesting characteristics. First, children who were not stunted had a large diverse group of variables closely associated with their physical development.

There is a broad range of general hypotheses on factors for reducing child stunting including farming practices, maize quality, micronutrient consumption, diarrheal occurrences, ORT usage, water access, breastfeeding, and prenatal health.

However, as the category of child stunting level dropped below the WHO defined stunting threshold (-2), the number of nodes in the first and second level dropped (23 to 9). The variables identified in the models for the stunted and extremely stunted children were very similar with only a slightly different structure. The similarity in model structure potentially suggests these variables are consistent in their effect on child growth. Interestingly, all of these variables are also present in the non-stunted child's model. This could suggest that not only are the identified variables in Figure 4 and 5 important, but to achieve improvements in child stunting the missing variables from Figure 3 should be considered.

4.3. Age-Specific Stunting Factors

In the age specific stunting models, several interesting trends were identified by the tool that warrant further investigation. First, animal pens were identified as strongly associated with the zhaz in the first two quantiles. Recent work has found links between farm animals fecal matter and the transmission of diseases [22]. Common in Guatemala, chickens and other farm animals are allowed to roam freely both near and inside the household. As children are yet to be walking between 0 and 12 months of age, this potential transmission route could play a significant role in a child's physical development.

Another emerging concern among health practitioners within Guatemala is the presence of mycotoxins in the maize supply and its effects on child growth [23]. Several organizations and academic institutions are investing resources to conduct research on improved storage techniques to reduce mycotoxin exposure. The models in this paper showed ‘improved maize storage’ as a first level factor for children under one year of age (0-6 and 6-12). This variable then drops to a fifth level factor for children between one and two. While the literature is sparse in linking mycotoxin exposure to child stunting this finding supports the continued efforts in identifying potential mechanistic links for younger children.

The role of nutrition can also be seen in the data. Models for children 0-6 and 7-12, identified only meat and cheese, respectively as being associated with zhaz. However, the children who were 13-18 or 19-24 months had general diet diversity as an important node in both models. Hygiene (soap and water available at the hand washing station) became a significant topic as the quantile shifted to children older than 12 months. This was similar for the language variable as well, which became a first level node for children 12 months or older. Nutrition, hygiene, and language have all been reported as significant factors in the health of children in Guatemala [24]. These findings support the literature and provide a base for multiple hypothesis testing of key relationships within these topics.

4.4. The Path Forward

As presented in this paper, weighted correlation network analysis could be a powerful asset to health information systems in Guatemala for understanding complex problems such as child stunting. These problems have major negative outcomes that affect many lives and have so far been resistant to effective intervention.

Next steps for the tool include expanding the analysis dataset and moving the software to a web interface. Aggregating both national and regional data sets would improve the accuracy of the models and help shed light on how they change over time with interventions. The outcomes and cofactors will be selectable so the user has ’switches’ they can use to manipulate models to look at different causal pathways. These switches may include municipality, language, gender, age, year, body mass index, wasting, and underweight. The introduction of other outcome variables could also be used in place of the zhaz score to explore contributing factors to other related health issues.

Further work also needs to be done for mathematical analysis of the transformed correlation network. A mathematical model should be created with parametric assumptions of the data to help choose the parameter β. Additional indices can be created to indicate how well each variable fits within its placement in the spanning tree this will ensure that users keep an open mind to other causal paths when looking at the trees, which present only the most significant. This tool could also provide academic researchers with a platform to use more advanced machine learning algorithms or regression tools to test hypotheses (as opposed to search for them).

The holistic analysis method and visual interface demonstrated here show viability for a powerful new health information system in Guatemala. Consistency with literature and ability to use many features of popular methods also ground this approach in traditional academic methods typically out of reach for end-users. The combination of novel methods with modern tools make this a good fit for solving major issues in analysis and communication that Guatemala health information systems currently face.

ACKNOWLEDGEMENTS

USAID for providing free access to their data, to the Schowalter Foundation, and to the people of Guatemala.

REFERENCES

(1) United Nations General Assembly, “Transforming our world: the 2030 Agenda for Sustainable Development,” pp. 1–35, 2015.

(2) World Health Organization, “Country Profile Indicators: Interpretation Guide,” Geneva, Switzerland, Tech. Rep., 2010.

(3) United Nations Children’s Emergency Fund, “Improving Child Nutrition: The achievable imperative for global progress,” UNICEF, Tech. Rep., 2013. [Online]. Available: http://www.unicef.org/gambia/Improving Child Nutrition - the achievable imperative for global progress.pdf.

(4) H. Alderman, J. Hodditnott, and B. Kinsey, “Long-term Consequences of Early Childhood Malnutrition,” Oxford Economic Papers, vol. 58, no. 168, pp. 450–474, 2006.

(5) K. G. Dewey and K. Begum, “Long-term consequences of stunting in early life,” Maternal and Child Nutrition, vol. 7, no. SUPPL. 3, pp. 5–18, 2011.

(6) A. W. Onyango, “Promoting healthy growth and preventing childhood stunting: A global challenge,” Maternal and Child Nutrition, vol. 9, no. S2, pp. 1–5, 2013.

(7) W. H. Organization, “Childhood Stunting : Context , Causes and Consequences WHO Conceptual framework,” vol. 9, no. September, 2013.

(8) USAID and Health Finance & Governance, “Guatemala Health System Assessment 2015,” Tech. Rep. August, 2015.

(9) WHO and PATH, “Optimize: Guatemala Report - Health Information Systems and Immunizations,” Tech. Rep. September, 2013. [Online]. Available: http://www.path.org/publications/files/TS opt guatemala rpt.pdf.

(10) WHO, “Developing Health Management Information Systems,” a Practical Guide for Developing Countries, p. 60, 2004.

(11) C. P. Wild, J. D. Miller, and J. D. Groopman, “Mycotoxin Control in Low- and Middle- Income Countries,” World Health Organization, Tech. Rep. 9, 2016.

(12) M. Bastian, S. Heymann, and M. Jacomy, “Gephi: An Open Source Software for Exploring and Manipulating Networks,” 2009. [Online]. Available: http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154.

(13) B. Zhang and S. Horvath, “Statistical Applications in Genetics and Molecular Biology A General Framework for Weighted Gene Co-Expression Network Analysis,” Statistical Applications in Genetics, vol. 4, no. 1, 2005.

(14) ICF International, “Baseline Study of the Title II Development Food Assistance Programs in Guatemala,” 2014.

(15) A. A. Hagberg, D. A. Schult, and P. J. Swart, “Exploring network structure, dynamics, and function using NetworkX,” Proceedings of the 7th Python in Science Conference (SciPy 2008), no. SciPy, pp. 11–15, 2008.

(16) J. D. Hunter, “Matplotlib: A 2D Graphics Environment,” Computing in Science {&} Engineering, vol. 9, no. 3, 2007.

(17) W. McKinney, “Data Structures for Statistical Computing in Python,” in Proceedings of the 9th Python in Science Conference, S. van der Walt and J. Millman, Eds., 2010, pp. 51–56.

(18) S. van der Walt, S. C. Colbert, and G. Varoquaux, “The NumPy Array: A Structure for Efficient Numerical Computation,” Computing in Science {&} Engineering, vol. 13, no. 2, 2011.

(19) N. Fenske, J. Burns, T. Hothorn, and E. a. Rehfuess, “Understanding child stunting in India: a comprehensive analysis of socio-economic, nutritional and environmental determinants using additive quantile regression.” PloS one, vol. 8, no. 11, p. e78692, jan 2013. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3817074&tool=pmcentrez&rendertype=abstract.

(20) D. Haile, M. Azage, T. Mola, and R. Rainey, “Exploring spatial variations and factors associated with childhood stunting in Ethiopia: spatial and multilevel analysis.” BMC pediatrics, vol. 16, no. 1, p. 49, 2016.

(21) B. S. Cade and B. R. Noon, “A gentle introduction to quantile regression for ecologists,” Frontiers in Ecology and the Environment, vol. 1, no. 8, pp. 412–420, 2003.

(22) M. N. N. Mbuya and J. H. Humphrey, “Preventing environmental enteric dysfunction through improved water, sanitation and hygiene: An opportunity for stunting reduction in developing countries,” Maternal and Child Nutrition, vol. 12, pp. 106–120, 2015.

(23) I. Fenton, “Problematizing the effect of rural-urban linkages on food security and malnutrition in Guatemala’s Western Highlands,” 2013.

(24) N. Goldman, A. R. Pebley, and M. Beckett, “Diffusion of ideas about personal hygiene and contamination in poor countries: Evidence from Guatemala,” Social Science and Medicine, vol. 52, no. 1, pp. 53–69, 2001.

II. WHAT CAUSES CHILDHOOD STUNTING AMONG CHILDREN OF SAN VICENTE, GUATEMALA: EMPLOYING COMPLIMENTARY, SYSTEM-ANALYSIS APPROACHES

To be submitted to the Journal of Environmental Health Perspectives

Lee E. Voth-Gaeddert1*, Matthew Stoker2, Devin Cornell3, Daniel B. Oerther1

1 Civil, Architectural, and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO, USA, lv6w3@mst.edu

2 Independent Researcher

3 Department of Sociology, University of California, Santa Barbara, Santa Barbara, CA, USA, dcornell@umail.ucsb.edu

*Corresponding Author

ABSTRACT

Background: Within the western hemisphere, Guatemala has the worst stunting rate with 49% of children under five years of age classified as stunted according to World Health Organization standards. The causes of this condition are not well known; therefore, it is unclear which interventions are the most cost effective to eliminate stunting. To begin to identify root causes, in this study, two different yet complimentary system-analysis approaches are used to analyze correlations among environmental and demographic variables, environmental enteric dysfunction (EED), and child height-for-age (stunting metric) in the community of San Vincente, Guatemala.

Methods: Based upon the literature and first-hand observations in the field, two descriptive models were constructed. The first model hypothesized relationships among EED and environmental and demographic variables, including: the presence of infrastructure to promote access to water, sanitation, and hygiene (WaSH). The second model hypothesized relationships among height-for-age and environmental and demographic variables, including: breastfeeding practices, the diversity of diet, prenatal health, aflatoxin burden, and child-mother interactions. The height-for-age model was also used to explore the confounding impact of EED on stunting. The descriptive models were analyzed using Network Analysis (NA) and Structural Equation Modeling (SEM) with data from two populations of children between the age of three months and five years. The first population (n=2,103) was drawn from the Food for Peace Baseline Survey conducted by the United States Agency for International Development (USAID) in 2012, and the second population (n=371) was drawn from an independent survey conducted by the San Vicente Health Center in 2016.

Findings: The results from the NA of the EED model confirmed water source, water treatment, and type of sanitation as important, and the results from the NA of the height-for-age model confirmed pathogen exposure, nutrition, and prenatal health as important. The results from the SEM of the EED model identified statistically significant correlations among EED with water source (-0.101, p=0.070) and type of water treatment (0.099, p=0.026). The results from the SEM of the height-for-age model identified statistically significant correlations among child height-for-age with prenatal health (0.121, p=0.074) and child-mother interaction (-0.091, p=0.079). Also, the SEM identified that aflatoxin burden (0.899, p=0.063) and child diet diversity (-0.136, p=0.092) were mediated by EED.

Interpretation: This is the first study to demonstrate complimentary system-analysis approaches to identify correlations among environmental and demographic variables, EED, and child height-for-age. Our approach supports the decision to use a multi-faceted intervention strategy to eliminate child stunting around San Vicente, and our results demonstrate an important tool that may be expanded to evaluate return on investment for strategies to eliminate child stunting throughout the western highlands of Guatemala.

Funding: Financial support was provided by the United States Peace Corps, the Conflict and Development Foundation of Texas A&M, and the Showalter Foundation

1. INTRODUCTION

Child stunting is defined as two standard deviations below the mean height-for-age as compared to the World Health Organization (WHO) growth chart (World Health Organization, 2010). Child stunting has been correlated with both acute and chronic health complications including increased morbidity as a child, increased risk of non-communicable diseases and obesity as an adult, and premature mortality (Alderman, Hodditnott, & Kinsey, 2006; Dewey & Begum, 2011). Children in Guatemala are among the most stunted in the western hemisphere and sixth worst in the world with rates of stunting at 49% (United Nations Children's Emergency Fund, 2013). Among rural Mayan communities in the western highlands of Guatemala, the rates of stunting are nearly 77% (United States Agency International Development, 2014). Child stunting is a difficult problem to address due to the high number of potentially associated causal variables. Hypothesized causal variables include micronutrient intake, caloric intake, breastfeeding practices, adequate water sources and treatment, proper sanitation, proper hygiene practices, and, recently proposed, exposure to fungal toxins (Black et al., 2013; Esrey, 1996; Solomons et al., 2014; Wild, Miller, & Groopman, 2016). In the current study, the confounding relationship among water, sanitation, and hygiene (WaSH) and fungal toxins are explored in relation to the putative role of environmental enteric dysfunction (EED) on stunting.

EED is an intestinal dysfunction identified by inflammation, villi blunting, and increased crypt depth (Ahmed et al., 2014). Chronic exposure to pathogens is hypothesized to be a causative factor for EED, and EED is believed to be more likely to occur among children living in environments lacking infrastructure to promote WaSH (Keusch et al., 2013). Members of the fungal genus, Aspergillus spp., biochemically produce aflatoxin B that has been identified as a group 1 carcinogen by the WHO due to negative effects on the liver (International Agency for Research on Cancer, 2006). Additionally, a recent review article published by the WHO hypothesized that exposure to high levels of aflatoxin B is a contributing factor to child stunting (Wild et al., 2016). However, due to the complex nature of the relationships among the variables potentially contributing to stunting, it is difficult to perform a holistic assessment to determine the most cost-effective intervention to prevent future stunting. Network analysis (NA) and structural equation modeling (SEM) provide two complimentary, system-analysis approaches for analyzing complex relationships. NA applies predetermined rules, in the form of algorithms, to describe the relationships among variables. NA often is applied to large data sets to identify putative correlations among input variables and specific outcomes (for example, child stunting) (Zhang & Horvath, 2005). SEM uses path analysis and factor analysis to test hypotheses about the relationships among directly observed and latent variables (Grace, 2006). Previously, we reported on the use of a two-step process combining an initial evaluation of large data sets with basic statistical techniques (Canonical Correlation Analysis, Latent Factor Regression, Malanobis Teguchi Strategy) followed by hypothesis testing with SEM and small data sets to evaluate the relationship among environmental variables and the occurrence of diarrhea in Brazil (L. E. Voth-Gaeddert, Divelbiss, & Oerther, 2015a; Voth-Gaeddert, Divelbiss, & Oerther, 2015b). In the current study, we expand our prior result using NA on a large data set and SEM with a small data set to analyze correlations among environmental and demographic variables, EED, and child height-for-age in the community of San Vincente, Guatemala. The combination of these methods demonstrates an important tool that may be expanded to evaluate return on investment for strategies to eliminate child stunting throughout the western highlands of Guatemala.

This study uses NA to mine a USAID dataset to identify environmental variables potentially correlated to child height-for-age, and then uses SEM to test factors impacting child height-for-age among children in the town of San Vicente, Guatemala. The SEMs specifically examine the questions: 1) does EED cause a reduction in child height-for-age?, 2) does aflatoxin exposure cause a reduction in child height-for-age?, and 3) does aflatoxin exposure cause a reduction in child height-for-age mediated by EED.

2. METHODS

2.1. Location and Data Collection

In this study, two datasets were analyzed; the first was the 2012 US Agency for International Development (USAID) Food for Peace Baseline Survey (United States Agency International Development, 2014) (n=2,103). The survey was administered orally to households in the local dialect in five departments (states) in 30 municipalities (counties) throughout Guatemala. The second data set was collected by the San Vicente Health Clinic located in San Vicente Buenabaj, Totonicapán, Guatemala (15 1’33.20N, 91 35’1.99W). Among both populations, the farming of maize was the primary source of income with only one harvest per year. The primary language varied among the USAID data set but included Quiché, Ixil, Mam, and Popti while the primary language in San Vicente was Quiché. The secondary language for the majority of participants in both data sets was Spanish. The elevation for San Vicente is 2,780 meters, with an average range of temperatures of 5.1C to 17C, and an annual rainfall of 1,310 mm. Among the locations covered by the USAID survey, elevations varied between 1,600 and 3,100 meters, temperatures varied between 9.5C and 20.9C, and annual rainfall varied between 800 and 2,700 mm.

The methodology for data collection of the two data sets were similar but had two primary differences. First, for the USAID survey, all data was collected within the household through the use of a questionnaire (administered orally by a translator in the local dialect to the mother) and with direct collection of child anthropometric measurements (height, weight, and age following WHO guidelines; (World Health Organization, 2008)). For the San Vicente survey, a questionnaire was administered orally in the mother's local dialect in a semi-private facility on the side-line of a health assembly hosted by the local health center for mothers of children below five years of age. Direct collection of child anthropometric measurements were performed by healthcare providers (height, weight, and age following WHO guidelines; (World Health Organization, 2008)). Second, for the San Vicente survey, household observations were collected during a subsequent house visit which followed the health assembly.

The USAID data were obtained from the USAID Data Repository (United States Agency International Development, 2012) and children below five years of age with no missing data for the variables of height-for-age z-score, diarrheal occurrences, and breastfeeding practices were selected for analysis. The San Vicente data were obtained in de-identified format from the San Vicente Health Center and children below five years of age with no missing data for the variables of height-for-age z-scores and diarrheal occurrences were selected for analysis. Definitions of the variables utilized in NA are given in the supplementary material, and the definitions of the variables used in the SEM are shown in Table 1. Further information for the data collection methodology for the USAID survey are discussed in the baseline report (United States Agency International Development, 2014). Institutional Review Board exemption for the use of de-identified data was attained from Missouri University of Science and Technology.

Table 1. Variables and descriptions included in the structural equation models for height-for-age and EED.

Variable	Definition
Height-for-Age Model	
Child Stunting	Measure of child's height divided by age and standardized following WHO guidelines
# of times played w child yesterday	The times the child had intentional play time yesterday
Consumed Prenatal Folic Acid Supplements	Mother consumed folic acid vitamins during pregnancy
Number of Prenatal Visits to Clinic	The number of visits the mother made to the health facility during her pregnancy
Consumed Prenatal Vitamins	Mother consumed prenatal vitamins during pregnancy
Child Yellow Eyes	The mother reported the child having yellow coloring in the eyes
Child Unexplained Appetite or Weight Loss	The mother reported the child losing weight or appetite unexplainably
Child Had Headaches	The mother reported the child having problems with chronic headaches
Child Had Unexplained Swelling	The mother reported the child having unexplained swelling in any extremities of the body
Child Diet Diversity Score	A score generated based on a 24-hour food recall following USAID guidelines
Child Diarrhea in Past Two Days	The mother reported the child having diarrhea in the past two days
Child Had Stomach Pain	The mother reported the child having experienced stomach pain
Number of Diarrheal Occurrences of Child in Past Two Weeks	The mother reported the number of times the child had separate bouts of diarrhea within two weeks
EED Model	
Water Storage	The observed type of water storage in the household
Type of Sanitation Facility	The observed type of sanitation facility used by the household
Presence of Animals in Kitchen Area	The observed presence of animals in the kitchen
Presence of Flies in Kitchen Area	The observed presence of flies in the kitchen
Cleanliness of Kitchen	The observed presence of trash in the kitchen
Water Source for Household	The observed type of water source in the household
Water Treatment	The observed type of water treatment in the household

2.2. Statistical Analysis

A weighted correlation NA was applied to the USAID data. A shortest-path algorithm was used in the analysis which utilizes the correlations between all variables, but focused on minimizing the distance between all variables and the child height-for-age variable. All variables may only be connected to child height-for-age through a single path which can be a direct relationship or through several other variables. The algorithm decides how a variable will be connected to the child height-for-age variable by calculating several weighted summations of correlations (i.e. single paths) for each variable simultaneously. The combination of paths with the lowest combined value is then selected. Variables closest to center variable (directly connected) have the strongest direct correlation with child height-for-age. The result is a hierarchical tree, or spanning-tree, stemming from the child height-for-age variable. Python was used to apply this algorithm to the USAID data that, after sub-setting, contained 88 variables and n = 2,103 children. Two spanning-tree graphics were generated using Cytoscape; 1) centering on child height-for-age (ZHAZ) and 2) centering on EED (Diarrhea was used as this was the best metric available to represent EED). From the output graphics, the structure of the data could then be assessed to identify variables relevant to the hypothesized SEMs. Further detail on the algorithm can be found in Voth-Gaeddert et al. 2016 (Lee E Voth-Gaeddert & Cornell, 2016) and the Python code can be found on Github (Cornell & Voth-Gaeddert, 2016).

The relationships identified in the spanning-tree graphics from the network analysis were incorporated into the set of hypotheses in each SEM. Field observations and a literature review provided additional information to improve the hypothesized models. Furthermore, several experts, both nationally and locally, were consulted about the structure of the set of hypotheses in the SEMs (Dary, 2016, personal communication; Baudilio, 2016, personal communication).

SEM is a statistical technique that utilizes path analysis and factor analysis to assess multiple hypotheses simultaneously. Factor analysis statistically determines the value of a hypothesized latent variable from a set of 'manifested' observable indicator variables (analogous to symptoms a doctor would look for to identify an underlying disease). Path analysis then utilizes the data driven covariance matrix of the latent and observable variables to assess their fit to the hypothesized covariance matrix generated from the hypothesized SEM (does the data match the model?). Path analysis is then able to account for mediating variables (an independent variable affecting a dependent variable through a mediating variable). Once specified, a SEM can be analyzed in two steps; first the data are compared to the hypothesized measurement model which includes only the latent variables and their indicator variables. Second, if the data fit the measurement model, assessed via four model fit metrics, the data is then compared to all hypotheses in the SEM. If the data also show good fit to the full SEM, the within-model parameter estimates are then assessed. Parameter estimates are given in both standardized and unstandardized format and are interpreted in the same way as a regression analysis. Model fit metrics include Chi-Square (RMSEA; $p>0.05$), Root Mean Square Error of

Approximation (RMSEA; <0.08), Confirmatory Fit Index (CFI; >0.90), and Tucker-Lewis Index (TLI; >0.90). This study utilized this methodology for both a child height-for-age model and an EED model. The Lavaan package in R 3.3.2 was utilized for the SEM analysis and further reading on SEM can be found in Grace 2006 (Grace, 2006).

3. RESULTS

3.1. Descriptive Statistics

Table 2 displays descriptive statistics for both the USAID and San Vicente datasets. The USAID data had 2,103 children, of which 48% were male and 52% were female. The mean age of all children was 29.0 months. The mean height-for-age level was -2.47 standard deviations, and 30% of mothers reported their child having had diarrhea in the past two weeks.

The San Vicente data had 372 children, 48% males and 52% females with the mean age of all children at 29.4 months. The mean height-for-age level was -2.56 standard deviations and 20% of mothers reported their child having had diarrhea in the past two weeks.

Table 2. Descriptive statistics for the USAID and San Vicente datasets.

Variable	USAID Dataset	San Vicente Dataset
Sample size	2,103 children	372 children
Girls	52%	52%
Boys	48%	48%
Mean age of child	29.0 months	29.4 months
Mean height-for-age z-score	-2.47 standard deviations	-2.56 standard deviations
Mean diarrheal prevalence	30%	20%

3.2. Stunting Network Analysis

The output for the network analysis of the height-for-age model (labeled ‘Child height-for-age’) was a spanning tree. Figure 1 depicts the variables correlated with Child height-for-age up to the third variable for simplicity. The results included three primary branches from Child height-for-age with three primary topical categories. The first category was pathogen exposure; soap was present at the hand washing station (Soap at

Washing Station), water was present at the hand washing station (Water at Washing Station), water was available at the water source (Water Available), child had diarrhea within the past two weeks (Child Had Diarrhea), and the mother had used oral rehydration therapy (ORT) on the child (ORT Used by Mother). The second category was micronutrient and caloric intake; the child is/was exclusively breastfed for the first six months of life (Child Exclusively Breastfed), child was breastfed up to second birthday (Child Breastfed), and the diet diversity of the child (Child Diet Diversity Score). The final category was prenatal health; the height of the mother (Mother's Height), the weight of the mother (Mother's Weight), and if the mother was currently pregnant (Mother Currently Pregnant). The final remaining variable was the age of the child (Age of Child).

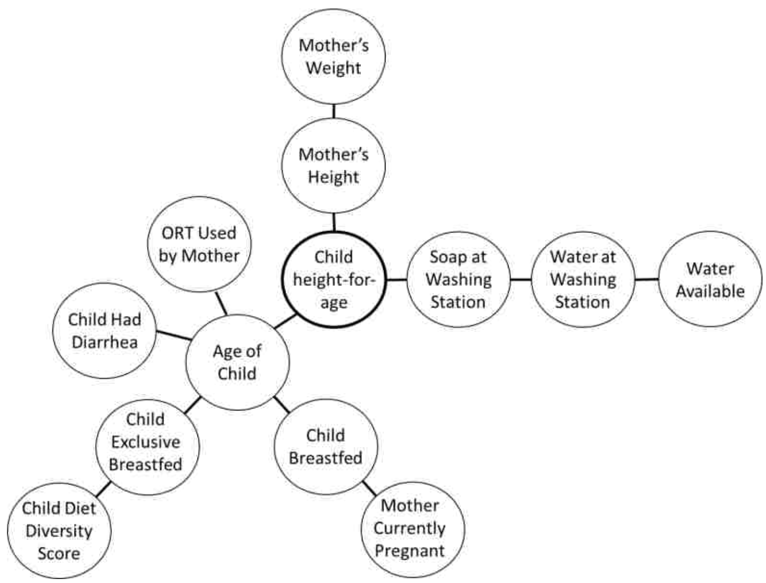


Figure 1. Spanning tree of height-for-age model modified from Voth-Gaeddert et al 2016 (Voth-Gaeddert & Cornell, 2016). Nodes are variables centered around 'Child height-for-age'. Lines are correlations selected by the algorithm as part of the shortest path of correlations to the child height-for-age variable for a given variable.

3.3. Diarrhea Network Analysis

The network analysis for the EED model centered the spanning tree around the variable Diarrhea. Figure 2 depicts the correlated variables with Diarrhea up to the third variable for simplicity. The results included three primary branches from Diarrhea and three primary topical categories. The first category included variables related to water availability; water not available at source in past month (No Water Available), water available at the handwashing station (Water at Washing Station), and water is currently available at water source (Water Available). The second category was Sanitation and included households sharing the sanitation facility (House Shared Sanitation). The final category included variables associated with removal of pathogens; soap present at the hand washing station (Soap at Washing Station) and the type of water treatment used in the household (Type of Water Treatment). The remaining variable was the age of the child (Age of Child).

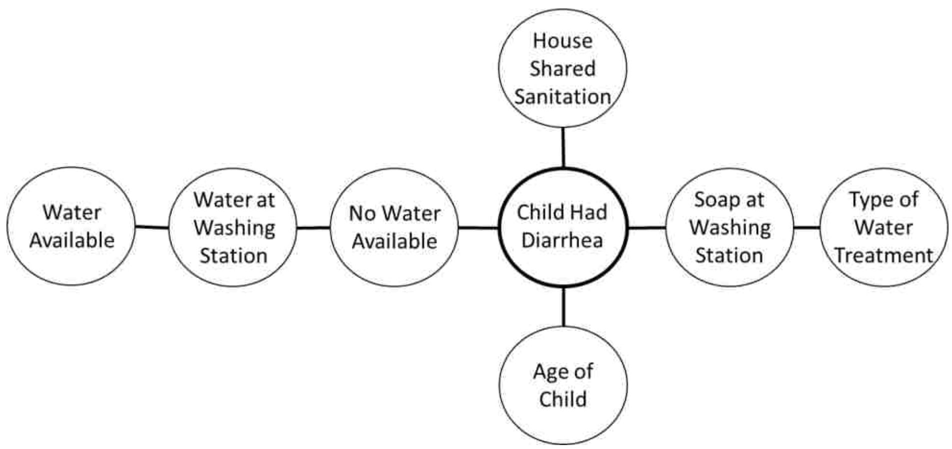


Figure 2. Spanning tree of EED model using child diarrhea modified from Voth-Gaeddert et al 2016 (Voth-Gaeddert & Cornell, 2016). Nodes are variables centered around 'Child Had Diarrhea'. Lines are correlations selected by the algorithm as part of the shortest path of correlations to the child had diarrhea variable for a given variable.

3.4. SEM of Child Height-for-Age Model

Combining the results from the network analysis, the literature review, and field observations, two SEMs were constructed for height-for-age and for EED. For the SEM of the height-for-age model there were three hypothesized latent variables - prenatal health, child aflatoxin burden, and EED – that made up the measurement model. The data showed good fit to the measurement model in all four measures of model fit providing justification for analyzing the full model. Subsequently, the data showed good fit to the full hypothesized SEM and yielded all four model fit tests successful (Chi-square: 81.086, p=0.100; RMSEA: 0.025 (CI: 0.000 – 0.043); Robust CFI: 0.968; Robust TLI: 0.956). The parameter estimates within the model could then be analyzed.

Figure 3 displays the results of the final SEM of the height-for-age model. Child height-for-age was regressed on by five variables; three had correlations below a 50% significance level and one had a statistically significant correlation below the 10% level. The observable variable, number of times child played yesterday (Child Played), was significant at a 10% level with a standardized parameter estimate of -0.092 (p=0.076). Additionally, the latent variable Prenatal Health was significant at a 15% level with a standardized parameter estimate of 0.151 (p=0.102). When the Prenatal Health variable was computed as a composite variable (as opposed to a latent variable) the correlation with EED became significant at a 5% level with a standardized parameter estimate of 0.121 (p=0.028). Neither EED nor Child Diet Diversity Score had statistically significant correlations with child height-for-age. Furthermore, three variables were regressed on by the mediating variable EED, none of which were statistically significant at a 10% level.

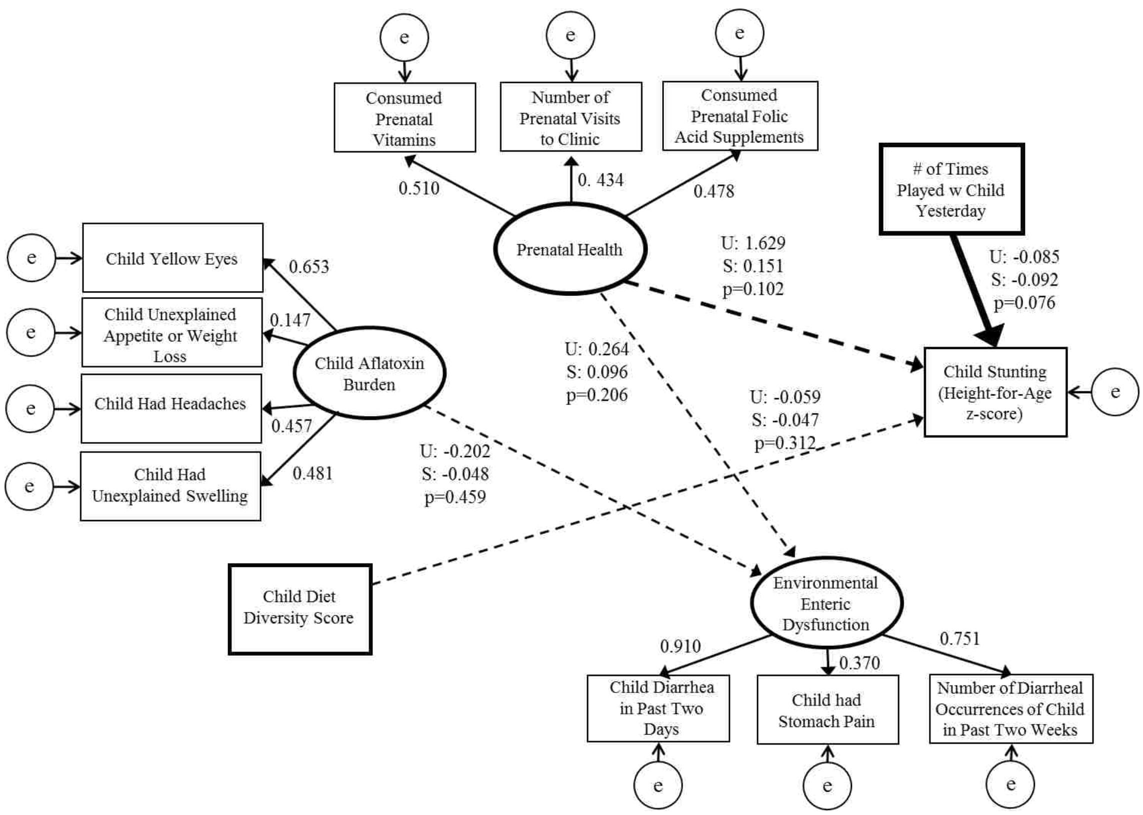


Figure 3. Final structural equation model of height-for-age model. DWLS robust estimator used; Chi-square: 81.086, p=0.100; RMSEA: 0.025 (CI: 0.000 – 0.043); Robust CFI: 0.968; Robust TLI: 0.956. Arrows are hypothesized direction, rectangles are observed variables, ovals are latent variables, ‘e’ are error. S = standardized parameter estimate, U = unstandardized parameter estimate, p = statistically significant level.

3.5. SEM of EED Model

The SEM for the EED model had two hypothesized latent variables – Food Preparation Habits and EED – which were first tested separately in a measurement model. Poor initial fit of the data to the model prompted the review of the model output statistics (the residual covariance matrix and modification indices). From this review, the 'Kitchen in a separate room' indicator variable of the Food Preparation latent variable was identified as the cause of the misfit. Further field observations and informal interviews were conducted with local mothers which identified that because the kitchen often was the primary family gathering place, food preparation was not correlated with structural investments in kitchens. Based on the confirmed discrepancy between the hypothesized model and the realities on the ground, this indicator variable was removed and the measurement model retested. Showing adequate fit in the measurement model (Chi-square: 8.677, p=0.370; Robust RMSEA: 0.013 (CI: 0.000 – 0.055); Robust CFI: 0.998; Robust TLI: 0.997), the full model could then be analyzed. Again, the initial fit of the data to the full model was poor; however, the model output statistics suggested a problem in the hygiene variable. Comparing raw data sets from this study and those of previous data collection campaigns conducted by the San Vicente Health Center, the presence of soap (the indicator used for Hygiene) was found to be above 95% among local households. With such high coverage, the variance within this variable was minimal and reduced the probability of identifying a correlation between other variables. With the removal of the hypothesized correlations with the Hygiene variable in the SEM, the full model showed good fit (Chi-square: 37.173, p=0.056; Robust RMSEA: 0.030 (CI: 0.000 – 0.049); Robust CFI: 0.981; Robust TLI: 0.967) and prompted the analysis of the standardized parameter estimates.

Figure 4 depicts the final result, including unstandardized and standardized parameter estimates, for the SEM of the EED model. Five variables were regressed directly on EED; three had significance levels below 50% and one had a significance level below 5%. Water Treatment had a parameter estimate of -0.115 (p=0.013) and was statistically significant at the 5% level. Water Source was statistically significant at the 15% level with a parameter estimate of 0.098 (p=0.127). Food Preparation Habits had a parameter estimate of -0.088 but was not statistically significant (not shown).

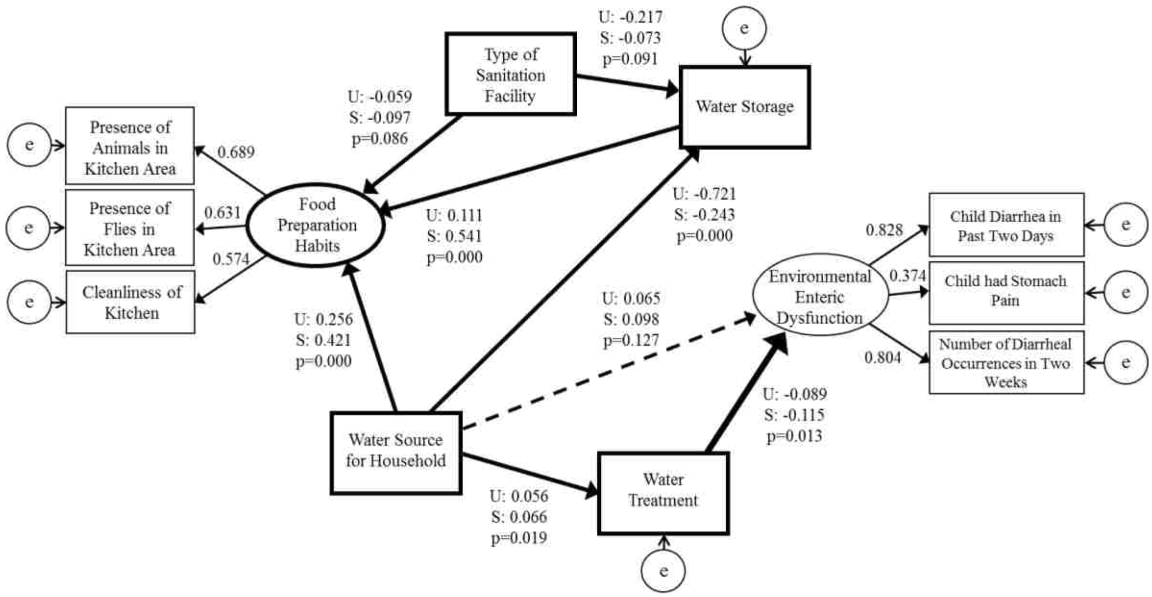


Figure 4. Final structural equation model of EED model. DWLS robust estimator used; Chi-square: 37.173, p=0.056; Robust RMSEA: 0.030 (CI: 0.000 – 0.049); Robust CFI: 0.981; Robust TLI: 0.967. Arrows are hypothesized direction, rectangles are observed variables, ovals are latent variables, 'e' are error. Solid arrows are confirmed statistically significant correlations at a 10%, dashed arrows are correlations important to the overall SEM but not significant at a 10% level. Size added for emphasis. S = standardized parameter estimate, U = unstandardized parameter estimate, p = statistically significant level.

Correlations with mediating variables included Water Treatment, regressed on Water Source; Water Storage regressed on Water Source and Sanitation Facility; and Food Preparation Habits, regressed on Water Source, Water Storage, and Sanitation Facility. Water Treatment had a statistically significant correlation at the 5% level with Water Source with a parameter estimate of 0.066 (p=0.019). Water Storage had a statistically significant correlation at the 0.1% level with Water Source with a parameter estimate of -0.243 (p=0.000). Water Storage also had a statistically significant relationship at the 10% level with Sanitation Facility with a parameter estimate of -0.073 (p=0.091). Finally, Food Preparation Habits had a statistically significant relationship at the 1% level with Water Source and Water Storage with parameter estimates of 0.421 (p=0.000) and 0.541 (p=0.000), respectively. Food Preparation Habits also had a statistically significant correlation with Sanitation Facility at the 10% level with a parameter estimate of -0.097 (p=0.086).

4. DISCUSSION

4.1. Network Analysis

The network analysis of the height-for-age model identified three categories; pathogen exposure, nutrition, and prenatal health. The pathogen exposure category consisted of three variables associated with pathogen transmission (HygSoap, HygWater, WaterAvai.) and two variables associated with potential pathogen exposure outcomes (Diarrhea and ORTuse). The United Nations Children's Fund (UNICEF), the WHO, and scholarly literature have identified pathogen exposure as a critical part of the general stunting model (Stewart, Iannotti, Dewey, Michaelsen, & Onyango, 2013; United Nations Children's Emergency Fund, 2013). The findings support the incorporation of WaSH or EED models in child stunting analyses. The presence of nutrition variables in the network analysis confirmed previous work on nutrition and stunting, specifically identifying diet diversity and breastfeeding practices as important (Georgieff, 2007; Rivera, Hotz, Gonzalez-Cossio, Neufeld, & Garcia-Guerra, 2003; Shugart, 2016). Finally, several prenatal health variables were identified and included potential multigenerational effects (mother's height and weight) and pregnancy status (mother is currently pregnant). Both factors have previously been shown to correlate with child stunting (Abuya, Ciera, & Kimani-Murage, 2012; Addo et al., 2013; Dewey & Cohen, 2007; Gipson, Koenig, & Hindin, 2012; Özaltin, Hill, & Subramanian, 2010). The data set did have limitations due to the types of questions in the areas of education, pregnancy health, and aflatoxin exposure.

The network analysis of the EED model (using diarrhea as a proxy) included three categories; water availability, sanitation, and pathogen removal. The variables in the category of water availability included water at the handwashing station, water at the house, and no water available from the most common water source for the household. Hunter et al 2010 review the implications of increased water stress on households, noting its direct and indirect relationship with pathogen exposure and diarrheal occurrences (Hunter, MacDonald, & Carter, 2010). Sanitation, specifically, households who shared a sanitation facility with another household, was identified as a separate branch correlated to diarrheal occurrences. Extensive previous research has reported sanitation-related factors as key potential barriers for the transmission of diarrheal diseases (Baker et al., 2016). Finally, several variables related to blocking or removing pathogens were identified and included having soap at the handwashing station, the type of water treatment device owned, and if the family utilized the water treatment device. Soap and improved water treatment devices have been reported as effective ways to reduce potential exposure to diarrheal diseases (Goldman, Pebley, & Beckett, 2001; Moll, McElroy, Sabogal, Corrales, & Gelting, 2007; Reller et al., 2003; Rosa, Miller, & Clasen, 2010; Stauber, Ortiz, Loomis, & Sobsey, 2009).

These results were augmented with a literature review and field observations to hypothesize two SEMs. The USAID data analyzed in the NA was collected from a geographically wider population as compared to the San Vicente data. However, many of the environmental challenges faced by residences across the western highlands of Guatemala are comparable (United States Agency International Development, 2014), thereby supporting the applicability of the NA results to the SEM hypotheses.

4.2. SEM for the Height-for-Age Model

The results of the SEM for the height-for-age model showed that two variables had statistically significant correlations with child height-for-age among children in San Vicente; Child Play and Prenatal Health. The variable Child Play was an observable variable that was negatively associated with child height-for-age. This meant that an increase in the number of times the child played, as reported by the mother, was associated with a decrease in child height-for-age. This finding was counter to the original hypothesis; however, during subsequent field observations mothers were observed permitting their children to play in potentially unsanitary conditions. The original intent of this variable was to capture potential hormone stimulation from the child-mother interaction and immobility of the child, but the data suggest it may have captured an additional pathogen transmission route instead. Ngure et al (2013) and Kolahi et al (2008) found that the cleanliness of a child’s play area was correlated with diarrheal occurrences (Kolahi, Nabavi, & Sohrabi, 2008; Ngure et al., 2013). Additionally, Voth-Gaeddert et al. (2016) found that in the western highlands of Guatemala having a fenced-in area for animals was correlated with the growth of the child within the first year of life, potentially due to the decreased probability of pathogen transmission via animal fecal matter (Voth-Gaeddert & Cornell, 2016).

While the prenatal health latent variable was close to being significant at a 10% level (p=0.102), the prenatal health composite variable was significant at a 5% (p=0.028) with child height-for-age. The prenatal health variable included three indicator variables; two questions based on vitamin supplement consumption during pregnancy (prenatal and folic acid) and one question on the number of visits to the health center during pregnancy. The data suggest that this factor was positively associated with the height-for-age z-score of a child. This finding was supported by informal interviews with local health facility staff. Additionally, the height-for-age z-score for children in the western highlands of Guatemala at birth has been reported at less than -1.00 (Solomons et al., 2014), suggesting that the prenatal period is critical for child growth. Finally, both EED and child diet diversity had statistically insignificant parameter estimates with child height-for-age, however, both contributed to the adequate fit of the data to the model and have been statistically significant with child stunting in previous research (Checkley et al., 2008; Georgieff, 2007), prompting further investigation.

The hypothesized mediating variable within the SEM for the height-for-age model was the EED latent variable. The EED latent variable was represented/manifested by three indicator variables; child had diarrhea in past two days, child had stomach pain in past two weeks, and number of diarrheal occurrences in the past two weeks. Previous studies have reported correlations between EED and diarrheal occurrences, chronic diarrhea, and intestinal sensitivity supporting the use of each (Korpe & Petri, 2012; Viswanathan, Hodges, & Hecht, 2009). However, the data did not show significant correlations between EED and either Child Aflatoxin Burden or Prenatal Health. Mapesa et al (2016) and Smith et al (2012) have proposed hypotheses for the mechanistic pathways in the association of aflatoxin and EED. Voth-Gaeddert et al (2017) recently reported a correlation between putative aflatoxin exposure of children in Guatemala and the four symptom-based indicator variables used in the Child Aflatoxin Burden latent variable in this study. However, they found a negative relationship between putative aflatoxin exposure and EED. Finally, the Prenatal Health latent variable was not significantly correlated with EED but was important to the overall fit of the model.

Matamoros et al 2013 discuss how intestinal health can be potentially affected by factors associated with healthy prenatal practices.

The boundaries set between the two SEMs were established based on the type of mechanistic pathway in which each variable impacted EED. Variables included in the SEM for the EED model were hypothesized to effect EED via an increased exposure to pathogens. Variables acting on EED in the SEM for the height-for-age model were hypothesized to effect EED through non-pathogenic mechanistic exposure pathways. Aflatoxin exposure has been hypothesized to affect the immune system by overstimulating cytokines and potentially causing inflammation and reduced absorptive capacity (Wild et al., 2016). Prenatal health has been hypothesized to affect the development of the intestinal microbiome and functionality later in life (Matamoros et al., 2013).

4.3. SEM for the EED Model

In the SEM for the EED model, two variables had statistically significant correlations at a 15% level with the EED latent variable. The EED latent variable in the SEM for the EED model included the three indicator variables previously used in the EED latent variable in the SEM for the height-for-age model. The two variables statistically correlated with EED were Water Source and Water Treatment. Water Treatment was negatively correlated at a 5% level with EED meaning that as the method of water treatment improved the level of EED decreased. This finding supported the original hypothesis and confirmed previous research on this relationship. Zwane et al conducted a review of the literature and found hygiene and point-of-use water treatment devices were effective in reducing intestinal disease exposure (Zwane & Kremer, 2007). Water Source had a positive correlation with EED meaning that as the quality of the water source decreased the level of EED decreased. This was contrary to the original hypothesis and previous research. Further investigation of the raw data revealed 81.3% of participants reported utilizing the community water distribution system, while 14.0% reported utilizing a faucet inside their house. This suggested that households who reported utilizing a faucet in the house, as opposed to the option of answering the community water distribution system, had children with less intestinal dysfunction (diarrhea/discomfort). The latent variable Food Preparation Habits had statistically insignificant correlation with EED, but contributed to the overall fit of the model and has been shown to be important in previous research (Agustina et al., 2013), prompting further study.

Three mediating variables in the SEM for the EED model included Water Treatment, Water Storage, and Food Preparation Habits. Water Source, while having a 'direct' correlation with EED also had an 'indirect' correlation via the mediation of Water Treatment. Water Source had a positive correlation with Water Treatment meaning that as the water source improved the water treatment technique also improved. This supported the original hypothesis and suggests that possible secondary benefits may accrue if the household's water source is improved. Water Source and Sanitation Facility had negative correlations with Water Storage suggesting that as either the water source or sanitation facility improved, water storage was worse. These findings were counter to the original hypotheses. Households could have possibly felt the protection provided by an improved water source or sanitation facility would be sufficient. Finally, Water Source and Water Storage had a positive correlation with Food Preparation Habits while Sanitation Facility had a negative correlation. Therefore, as the water source or water storage improved, food preparation habits improved as well; however, when sanitation facilities improved food preparation habits worsened. Zwane et al (2007) discuss the potential complexities of WaSH infrastructure and diarrheal occurrences in developing countries (Zwane & Kremer, 2007).

This study analyzed the factors hypothesized to be correlated with child height-for-age and EED in the town of San Vicente, Guatemala. Two models were developed and tested utilizing two system-analysis approaches; NA and SEM. Results confirmed the hypothesis that for children in San Vicente Child Play and Prenatal Health were correlated with child height-for-age. Additionally, the type of water treatment and type of water source were identified as significant for EED. The sum of these results suggests a complex reality within the environmental and demographic based factors hypothesized to affect child stunting. Practitioners must understand these complex realities on the ground and utilize the appropriate tools for identifying effective interventions.

REFERENCES

Abuya, B. A., Ciera, J., & Kimani-Murage, E. (2012). Effect of mother's education on child's nutritional status in the slums of Nairobi. BMC Pediatrics, 12(1), 80. https://doi.org/10.1186/1471-2431-12-80.

Addo, O. Y., Stein, A. D., Fall, C. H., Gigante, D. P., Guntupalli, A. M., Horta, B. L., … Martorell, R. (2013). Maternal height and child growth patterns. Journal of Pediatrics, 163(2), 549–554.e1. https://doi.org/10.1016/j.jpeds.2013.02.002.

Agustina, R., Sari, T. P., Satroamidjojo, S., Bovee-Oudenhoven, I. M. J., Feskens, E. J. M., & Kok, F. J. (2013). Association of food-hygiene practices and diarrhea prevalence among Indonesian young children from low socioeconomic urban areas. BMC Public Health, 13, 977. https://doi.org/10.1186/1471-2458-13-977.

Ahmed, T., Auble, D., Berkley, J. a, Black, R., Ahern, P. P., Hossain, M., … Gordon, J. I. (2014). An evolving perspective about the origins of childhood undernutrition and nutritional interventions that includes the gut microbiome. Annals of the New York Academy of Sciences, 1332, 22–38. https://doi.org/10.1111/nyas.12487.

Alderman, H., Hodditnott, J., & Kinsey, B. (2006). Long-term Consequences of Early Childhood Malnutrition. Oxford Economic Papers, 58(168), 450–474.

Baker, K. K., O'Reilly, C. E., Levine, M. M., Kotloff, K. L., Nataro, J. P., Ayers, T. L., … Mintz, E. D. (2016). Sanitation and Hygiene-Specific Risk Factors for Moderate-to-Severe Diarrhea in Young Children in the Global Enteric Multicenter Study, 2007-2011: Case-Control Study. PLoS Medicine, 13(5), 1–19. https://doi.org/10.1371/journal.pmed.1002010.

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. a, Christian, P., de Onis, M., … Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet, 382(9890), 427–51. https://doi.org/10.1016/S0140-6736(13)60937-X.

Checkley, W., Buckley, G., Gilman, R. H., Assis, A. M., Guerrant, R. L., Morris, S. S., … Black, R. E. (2008). Multi-country analysis of the effects of diarrhoea on childhood stunting. International Journal of Epidemiology, 37(4), 816–830. https://doi.org/10.1093/ije/dyn099.

Cornell, D., & Voth-Gaeddert, L. E. (2016). Github: Weight Correlation Network Analysis Code. Retrieved from https://github.com/devincornell/weighted-correlation-network-analysis.

Dewey, K. G., & Begum, K. (2011). Long-term consequences of stunting in early life. Maternal and Child Nutrition, 7(SUPPL. 3), 5–18. https://doi.org/10.1111/j.1740-8709.2011.00349.x.

Dewey, K. G., & Cohen, R. J. (2007). Does birth spacing affect maternal or child nutritional status? A systematic literature review. Maternal and Child Nutrition, 3(3), 151–173. https://doi.org/10.1111/j.1740-8709.2007.00092.x.

Esrey, S. A. (1996). Water , Waste , and Well-Being : A Multicountry Study. American Journal of Epidemiology, 143(6), 608–623.

Georgieff, M. K. (2007). Nutrition and the developing brain: Nutrient priorities and measurement. American Journal of Clinical Nutrition, 85 (suppl), 614S–620S. https://doi.org/85/2/614S [pii].

Gipson, J. D., Koenig, M. A., & Hindin, M. J. (2012). The Effects of Unintended Pregnancy on Infant, Child, and Parental Health: A Review of the Literature. Studies in Family Planning, 39(1), 18–38. Retrieved from http://onlinelibrary.wiley.com.ezproxy.welch.jhmi.edu/doi/10.1111/j.1728-4465.2008.00148.x/abstract http://onlinelibrary.wiley.com.ezproxy.welch.jhmi.edu/store/10.1111/j.1728-4465.2008.00148.x/asset/j.1728-4465.2008.00148.x.pdf?v=1&t=h5x4hk5x&s=dedcf3d5bf.

Goldman, N., Pebley, A. R., & Beckett, M. (2001). Diffusion of ideas about personal hygiene and contamination in poor countries: Evidence from Guatemala. Social Science and Medicine, 52(1), 53–69. https://doi.org/10.1016/S0277-9536(00)00122-2.

Grace, J. B. (2006). Structural Equation Modeling and Natural Systems (First). United Kingdom: Cambridge University Press.

Hunter, P. R., MacDonald, A. M., & Carter, R. C. (2010). Water Supply and Health. PLoS Medicine, 7(11). https://doi.org/10.1371/journal.pmed.1000361.

International Agency for Research on Cancer. (2006). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Aflatoxins.

Keusch, G. T., Rosenberg, I. H., Denno, D. M., Duggan, C., Guerrant, R. L., Lavery, J. V., … Brewer, T. (2013). Implications of acquired environmental enteric dysfunction for growth and stunting in infants and children living in low-and middle-income countries. Food and Nutrition Bulletin, 34(3), 357–364.

Kolahi, A. A., Nabavi, M., & Sohrabi, M. R. (2008). Epidemiology of Acute Diarrheal Diseases Among Children Under 5 Years of Age in Tehran, Iran. Iranian Journal of Clinical Infectious Diseases, 3(4), 193–198.

Korpe, P. S., & Petri, W. a. (2012). Environmental enteropathy: Critical implications of a poorly understood condition. Trends in Molecular Medicine, 18(6), 328–336. https://doi.org/10.1016/j.molmed.2012.04.007.

Mapesa, J. O., Maxwell, A. L., & Ryan, E. P. (2016). An Exposome Perspective on Environmental Enteric Dysfunction. Environmental Health Perspectives, 124(8), 1121–1126.

Matamoros, S., Gras-Leguen, C., Le Vacon, F., Potel, G., & De La Cochetiere, M. F. (2013). Development of intestinal microbiota in infants and its impact on health. Trends in Microbiology, 21(4), 167–173. https://doi.org/10.1016/j.tim.2012.12.001.

Moll, D. M., McElroy, R. H., Sabogal, R., Corrales, L. F., & Gelting, R. J. (2007). Health impact of water and sanitation infrastructure reconstruction programmes in eight Central American communities affected by Hurricane Mitch. Journal of Water and Health, 5(1), 51–65. https://doi.org/10.2166/wh.2006.047.

Ngure, F. M., Humphrey, J. H., Mbuya, M. N. N., Majo, F., Mutasa, K., Govha, M., … Stoltzfus, R. J. (2013). Formative research on hygiene behaviors and geophagy among infants and young children and implications of exposure to fecal bacteria. American Journal of Tropical Medicine and Hygiene, 89(4), 709–716. https://doi.org/10.4269/ajtmh.12-0568.

Özaltin, E., Hill, K., & Subramanian, S. V. (2010). Association of Maternal Stature With Offspring Mortality, Underweight, and Stunting in Low- to Middle-Income Countries. Journal of the American Medical Association, 303(15), 1507–1516. https://doi.org/10.1001/jama.2010.450.

Reller, M. E., Mendoza, C. E., Lopez, M. B., Alvarez, M., Hoekstra, R. M., Olson, C. A., … Luby, S. P. (2003). A randomized controlled trial of household-based flocculant-disinfectant drinking water treatment for diarrhea prevention in rural Guatemala. American Journal of Tropical Medicine and Hygiene, 69(4), 411–419.

Rivera, J. a., Hotz, C., Gonzalez-Cossio, T., Neufeld, L. M., & Garcia-Guerra, A. (2003). The Effect of Micronutrient Deficiences on Child Growth: A Review of Results from Community-Based Supplementation Trials. Journal of Nutrition, 133, 4010S–4020S.

Rosa, G., Miller, L., & Clasen, T. (2010). Microbiological effectiveness of disinfecting water by boiling in rural Guatemala. American Journal of Tropical Medicine and Hygiene, 82(3), 473–477. https://doi.org/10.4269/ajtmh.2010.09-0320.

Shugart, B. J. (2016). MOTHER LODE Superhero sugars in breast milk make the newborn gut safe for beneficial bacteria. Science News, 185(1), 22–26.

Smith, L. E., Stoltzfus, R. J., & Prendergast, a. (2012). Food Chain Mycotoxin Exposure, Gut Health, and Impaired Growth: A Conceptual Framework. Advances in Nutrition: An International Review Journal, 3(4), 526–531. https://doi.org/10.3945/an.112.002188.

Solomons, N. W., Vossenaar, M., Chomat, A., Doak, C. M., Koski, K. G., & Scott, M. E. (2014). Stunting at birth: recognition of early-life linear growth failure in the western highlands of Guatemala. Public Health Nutrition, 18(10), 1737–1745.

Stauber, C. E., Ortiz, G. M., Loomis, D. P., & Sobsey, M. D. (2009). A randomized controlled trial of the concrete biosand filter and its impact on diarrheal disease in Bonao, Dominican Republic. American Journal of Tropical Medicine and Hygiene, 80(2), 286–293.

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 and Child Nutrition, 9(S2), 27–45. https://doi.org/10.1111/mcn.12088.

United Nations Children's Emergency Fund. (2013). Improving Child Nutrition: The achievable imperative for global progress. Retrieved from http://www.unicef.org/gambia/Improving_Child_Nutrition_-_the_achievable_imperative_for_global_progress.pdf.

United States Agency International Development. (2012). Development Data Library. Retrieved from https://www.usaid.gov/data.

United States Agency International Development. (2014). Baseline Study for the Title II Development Food Assistance Programs in Guatemala.

Viswanathan, V. K., Hodges, K., & Hecht, G. (2009). Enteric infection meets intestinal function: how bacterial pathogens cause diarrhoea. Nature Reviews Microbiology, 7(2), 110–119. https://doi.org/10.1016/j.surg.2006.10.010.Use.

Voth-Gaeddert, L. E., & Cornell, D. (2016). Improving health information systems in Guatemala using weighted correlation network analysis. IEEE Global Humanitarian Technology Conference, 686–693.

Voth-gaeddert, L. E., Divelbiss, D. W., & Oerther, D. B. (2015). Utilizing Structural Equation Modeling to Correlate Biosand Filter Performance and Occurrence of Diarrhea in the Village of Enseado do Aritapera in Para, Brazil. Water Science and Technology: Water Supply, 15(1), 164–172. https://doi.org/10.2166/ws.2014.101.

Voth-Gaeddert, L. E., Divelbiss, D. W., & Oerther, D. B. (2015). Utilizing structural equation modeling as an evaluation tool for critical parameters of the biosand filter in a pilot study in Para, Brazil. Water Science and Technology: Water Supply, 15(4), 845–851.

Voth-Gaeddert, L. E., Stoker, M., Torres, O., & Oerther, D. B. (n.d.). Analysis of Correlations between Aflatoxin B, Intestinal Function, and Child Height-for-Age Among Young Children in Guatemala. International Journal of Epidemiology.

Wild, C. P., Miller, J. D., & Groopman, J. D. (2016). Mycotoxin Control in Low- and Middle- Income Countries.

World Health Organization. (2008). Training Course on Child Growth Assessment. Geneva, Switzerland: WHO.

World Health Organization. (2010). Country Profile Indicators: Interpretation Guide. Nutrition Landacape Information System. Geneva, Switzerland. https://doi.org/10.1159/000362780.Interpretation.

Zhang, B., & Horvath, S. (2005). Statistical Applications in Genetics and Molecular Biology A General Framework for Weighted Gene Co-Expression Network Analysis A General Framework for Weighted Gene Co-Expression Network Analysis ∗. Statistical Applications in Genetics and Molecular Biology, 4(1). https://doi.org/10.2202/1544-6115.1128.

Zwane, A. P., & Kremer, M. (2007). What works in fighting diarrheal diseases in developing countries? A critical review. World Bank Research Observer, 22, 1–24. https://doi.org/10.1093/wbro/lkm002.

III. ANALYSIS OF CORRELATIONS AMONG AFLATOXIN B, ENTERIC DYSFUNCTION, AND CHILD HEIGHT-FOR-AGE AMONG YOUNG CHILDREN IN GUATEMALA

To be submitted to the International Journal of Epidemiology

Lee E. Voth-Gaeddert,1* Matthew Stoker,2 Olga Torres3 and Daniel B. Oerther1

1 Civil, Architectural, and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO, USA, lv6w3@mst.edu

2 Independent Researcher

3 Diagnósticos Molecular, Edificio Multimedica, Guatemala City, Guatemala

*corresponding author Lv6w3@mst.edu

ABSTRACT

Background: Recent research has reported a correlation between environmental enteric dysfunction (EED) and child height-for-age. Many factors may contribute to EED including the fungal toxin, aflatoxin B (AFB). This study reports correlations of 1) AFB exposure and potential symptoms of AFB and 2) AFB exposure and EED and height-for-age among children from San Vincente, Guatemala.

Methods: In October 2016 and February 2017, mothers with children ages four months to five years participated in health assemblies hosted by local health practitioners in San Vicente. A survey was orally administered in the local language to mothers and included a food recall, AFB related symptom questions, EED related symptom questions, and anthropometric measurements of the children. Subsequently after each assembly, house visits were conducted with the households of the mothers who attended the health assemblies. Samples of maize-to-be-consumed were collected from the households and a putative AFB consumption level was calculated for each child based on the level of AFB identified via the enzyme-linked immunosorbent assay test and the amount of consumed maize reported in the food recall. Two datasets were created, 1) data from all participants in the October 2016 health assemblies (n=320) and 2) data from participants who had attended both assemblies; October 2016 and February 2017 (n=120). The hypothesized correlations were tested with these datasets using the Kruskal-Wallis test, ordinal regression, factor analysis, and structural equation modeling (SEM).

Results: The mean putative AFB consumption level among children in October 2016 was 48.0 ng/kg of body weight. The putative AFB consumption level was significantly correlated with the October 2016 AFB symptoms variable (0.092, p=0.068). Furthermore, among participants who attended both health assemblies, the putative AFB consumption level in October 2016 was correlated at 1% level with AFB symptoms in October 2016 (0.123, p=0.026), at a 10% level with the change in AFB symptoms between 2016 and 2017 (0.107, n=0.099), and was not correlated with AFB symptoms in 2017. The putative AFB consumption level and AFB symptoms variable had a significant negative correlation with the EED symptoms variable (-0.093, p=0.036 and -0.133, p=0.006, respectively). The SEM analysis showed that there was a significant negative correlation between the putative AFB consumption level and EED symptoms variable (-0.080, p=0.030) and a significant negative correlation between the putative AFB consumption level and child height-for-age (-0.073, p=0.030). However, there was not a statistically significant relationship between EED and child height-for-age.

Conclusion: This is the first study to investigate the correlations between AFB exposure, EED symptoms, and child height-for-age in Guatemala. Based on the high exposure rates of AFB in Guatemala, further consideration should be given to the role of AFB exposure on child health.

Keywords: Aflatoxin B, Environmental Enteric Dysfunction, Height-for-Age, Factor Analysis, Structural Equation Modeling

KEY MESSAGES

- Putative AFB consumption levels among young children were statistically associated with four symptom questions of aflatoxin exposure
- Putative AFB consumption levels among young children were statistically associated with three symptom questions for environmental enteric dysfunction
- The relationship between putative AFB consumption levels and child height-for-age was mediated by environmental enteric dysfunction

1. INTRODUCTION

Malnutrition has been hypothesized to be an underlying contributing factor to 45% of all child deaths globally and is associated with both acute and chronic health problems.[1] Child height-for-age was selected as a global health indicator for child malnutrition by the United Nations General Assembly with the ratification of the Sustainable Development Goals.[2] Intestinal dysfunction has been reported to be negatively correlated with child height-for-age.[3] For children living in environments lacking WaSH infrastructure, chronic exposure to enteric pathogens can lead to a type of intestinal dysfunction named environmental enteric dysfunction (EED).[4] Conditions of EED include intestinal disturbances such as the blunting of villi, inflammation, and increased crypt depth which can lead to reduced absorptive capacity of the intestines. The majority of scholarly literature on EED investigates the effect of bacterial exposures; however, several recent review articles have reemphasized the negative effects due to toxic chemicals.[5] Specifically, Mapesa et al and Smith et al, have proposed that fungal toxins may be a contributor to EED and a set of mechanistic pathways have been hypothesized for mycotoxins on EED and child height-for-age.[5,6]

Aflatoxin B (AFB) is a carcinogenic type of mycotoxin and is produced by the fungi Aspregillus spp.[7] It is classified as a Group I carcinogen according to the International Agency for Research on Cancer (IARC).[8] The AFB strain is the most carcinogenic and is prevalent in a variety of crops including maize, sorghum, and groundnuts.[9] In 2016 the World Health Organization (WHO) published a review article citing evidence from six human studies from Africa and numerous animal studies on the potential links between aflatoxin exposure and reduced child height-for-age.[10] However, the current price of AFB biomarkers limit engagement from the research community resulting in unclear mechanistic pathways of AFB on EED and child height-for-age. Proposed effects of AFB on child health that are in common with EED include reduced zinc bioavailability, nutrient metabolism, protein synthesis, and damaged enterocytes.[10] Lizárraga-Paulín et al suggest that for children AFB exposure should be under 1 part per billion (ppb) in food.[11] Wild et al report on the disparities in the levels of AFB consumption between populations living in developed versus developing regions highlighting North America at 0-1 ng/kg of body weight and The Gambia at 4-113 ng/kg of body weight which is representative of many developing countries.[10] Additionally, the United States and European Union have set toxicity levels for imported maize at 20 ppb[12] and 5 ppb[13], respectively. In Guatemala, mean AFB levels in maize samples from local markets were found to be above US import limits in 11 of 24 departments and EU limits in 19 of 24 departments.[14] With high reported AFB levels and potentially negative health effects on children, AFB must be a priority in Guatemala.

As maize is a staple food among the people of Guatemala, AFB exposure to children is hypothesized to be high. In this study, data collected in October 2016 and February 2017 on children in San Vicente, Guatemala was used to test the hypothesized relationships between 1) putative AFB consumption levels from maize and AFB symptoms and 2) putative AFB consumption levels from maize, EED symptoms, and children's height-for-age.

2. METHODS

2.1. Location and Data Collection

The study site selected was a set of Mayan communities near the town of San Vicente Buenabaj, in the western highlands of Guatemala (15 1'33.20N, 91 35'1.99W). Communities lived among a mountain range with an average elevation of 2,780 meters and average range of temperatures of 5.1C to 17.0C. Farming of maize was the primary source of income while the primary language was Quiché and the secondary language was Spanish. The site had only one harvest season with the majority of households storing and consuming their own maize over the course of the year.

In October 2016, one month before harvest, and February 2017, two months following harvest, health assemblies were held by the local health center staff for mothers of children between three months and five years of age. Surveys were administered orally in the local dialect by local translators while anthropometric measurements were taken of the children by trained health professionals. The survey combined questions from the Demographic and Health Survey program[15], local health surveys, and AFB and EED symptom questions. Direct collection of anthropometric measurements and computations were conducted by trained nurses from the local health center following the WHO guidelines. House visits of the attending families were conducted one week after the health assemblies to collect household observational data and maize samples. The maize samples were collected from the stored maize allocated for consumption, immediately deposited into a paper bag, and sent to Guatemala City for analysis.

Two datasets were created from children with complete collected information. The first dataset included children who attended the October 2016 health assembly (n=320) while the second dataset was two time-points and included those children who attended both the October 2016 and February 2017 health assemblies (n=120). Institutional Review Board exemption for a chart review of information collected by the Health Center was attained from the Missouri University of Science and Technology (Missouri S&T) and the local San Vicente Health Center. All information was attained by the local Health Center under a licensed professional, de-identified, and subsequently analyzed by a team of researchers at Missouri S&T.

2.2. Aflatoxin Exposure Assessment

To assess putative aflatoxin exposure among children two types of measurements were used. First, maize samples, collected in accordance with Torres et al[16], were analyzed using the enzyme-linked immunosorbent assay (ELISA) test to obtain the amount of Aflatoxin B in a household's maize supply that was designated for consumption. Utilizing the 24-hour food recall portion of the survey (following US Agency for International Development guidelines[17]) a total amount of maize consumed in one day by the child was calculated. Multiplying the amount of AFB per gram of maize by the grams of maize consumed by the child in one day, an estimate of the average amount of AFB consumed in a single day by the child was computed. Finally, the AFB value was divided by the weight of the child to produce a comparable value across sampling ages. The name 'putative AFB consumption level' is used to identify this variable.

The second measurement was created from a set of questions on the survey administered to the mother and were based on potential symptoms related to aflatoxin exposure. These symptoms included yellow eyes, unexplained appetite or weight loss, body swelling, issues with urination, and chronic headaches.[18,19] These questions were asked of the child as well as any additional household members. As these symptoms were related to liver problems, one assumption made in the study was that if the liver was exposed to AFB the intestines were also exposed to similar levels of AFB. The name ‘AFB symptom latent’ is used to identify the combination of symptom-based questions when factor analysis is applied (see below) and ‘AFB symptom composite’ is used to identify these set of questions when the responses are summed.

2.3. Environmental Enteric Dysfunction Assessment

As EED is a broadly defined term associated with intestinal health, a set of symptom questions were given on the survey related to gastrointestinal problems of the child. These symptoms included; occurrence of diarrhea, rate of occurrence of diarrhea, rate of occurrence of dysentery, intestinal discomfort, and the most common illnesses within the household.[20] The name ‘EED symptom latent’ is used to identify the combination of questions when factor analysis is applied (see below) and ‘EED symptom composite’ is used to identify these set of questions when the responses are summed.

2.4. Statistical Analysis

Four statistical techniques were applied to the data to assess the significance between 1) the AFB consumption level and the AFB symptoms (latent and composite), 2) the AFB consumption level and the EED symptoms (latent and composite), and 3) the AFB consumption level, the EED symptom latent, and the child’s height-for-age. A cross sectional design was utilized for data from October 2016 and, where appropriate, a two time-point regression analysis was utilized for the two time-point dataset. The four techniques included Kruskal-Wallis rank sum test, ordinal regression, factor analysis, and structural equation modeling (SEM). Based on previous studies child food consumption and socio-economic status were controlled for in each model.[21] The statistical package R 3.3.2 was used for all analyses.

The Kruskal-Wallis rank sum test is used to assess the statistical significance between two variables with non-normal distributions in their data. The test was utilized to assess relationships between AFB consumption levels and the AFB symptoms composite, AFB consumption levels and the EED symptom composite, AFB consumption levels and child height-for-age, and the AFB symptoms composite and the EED symptom composite. McKnight et al provides an overview of the Kruskal-Wallis test.[22]

Ordinal regression is used to assess statistical correlations between an ordinal endogenous variable and its regressors. The test was utilized to assess the correlation between the EED composite and the AFB consumption level. Further information on ordinal regression can be found with Armstrong et al.[23]

Factor analysis is used to assess the latent structure of a set of variables hypothesized to manifest from the same source. As dichotomous, ordinal, and continuous variables were among the manifest variables in the analysis a robust diagonally-weighted least squares estimator was utilized to assess the latent factors.[24] Relationships analyzed included AFB consumption levels and the AFB symptom latent variable, AFB consumption levels and the EED symptom latent variable, and the AFB symptom latent variable and the EED symptom composite score. Graphically, for both factor analysis and SEM, the arrows depict the hypothesized directionality of effect, boxes are observable variables, and the ovals are the latent variable. Further reading on factor analysis can be found with Grace 2006.[25]

Finally, SEM is a technique that analyzes multiple hypotheses simultaneously, including potential mediating variables in a model. SEM combines factor analysis with path analysis to assess the relationship between three or more variables including observed, latent or composite. As with factor analysis, due to the presence of dichotomous and ordinal manifest variables in the SEM a robust diagonally-weighted least squares estimator was utilized. SEM was used to analyze the potential mediating effect of EED between AFB exposure and child height-for-age. Again, Grace 2006 provides further detail on the SEM methodology.[25]

3. RESULTS

3.1. Basic Statistics

There were 320 children between the ages of six months and five years in the October 2016 dataset and 120 children in the two time-point dataset. Table 1 displays descriptive statistics for both datasets. In the October 2016 dataset 49% were males and 51% were females with a mean age of 30.2 months. The mean height-for-age was -2.54

SD, the mean AFB consumption level was 48.0 ng/kg of body weight, and the mean prevalence of diarrhea for children within the previous two weeks reported by the mothers was 20.1%. For the two time-point dataset at time-point one (October 2016) 49% were males and 51% were females with a mean age of 30.7 months. The mean height-for-age was -2.66 SD, the mean AFB consumption level was 50.0 ng/kg of body weight and the mean prevalence of diarrhea for children within the previous two weeks reported by the mothers was 18%.

Table 1. Descriptive statistics for both the October 2016 and two time-point datasets.

Variable	October 2016 Dataset	Two Time-Point Dataset
Sample size	320 children	120 children
Girls	51%	51%
Boys	49%	49%
Mean age of child	30.2 months	30.7 months
Mean height-for-age z-score	-2.54 standard deviations	-2.66 standard deviations
AFB Consumption Level	48.0 ng/kg of body weight	50.0 ng/kg of body weight
Mean diarrheal prevalence	20.1%	18.0%

3.2. AFB Consumption Level vs AFB Symptoms

The first set of analyses assessed the hypothesized correlation between AFB consumption level and the AFB symptom-based questions, with both latent and composite variables. For the October 2016 dataset, the Kruskal-Wallis test did not confirm a significant correlation between the AFB consumption level and the AFB symptom composite (p=0.313). Figure 1 depicts the factor analysis (AFB symptom latent) results for October 2016. The factor analysis confirmed a statistically significant correlation at a 10% level between the AFB consumption level and the AFB symptom latent variable (0.092, p=0.068; X2 p=0.686, Robust RMSEA=0.000 (CI:0.000-0.039), Robust CFA=1.000, Robust TLI=1.154).

Additionally, the two time-point dataset was analyzed to assess the hypothesized correlation between the AFB consumption level and the AFB symptom latent as symptoms continued further into the future (see Figure 2). The Kruskal-Wallis test did not confirm a significant correlation between AFB consumption level and AFB symptom composite for any of the three measures. However, the factor analysis, see Figure 2, confirmed a correlation at the 5% level for symptoms in October 2016 (0.123, p=0.026), 10% for the change in symptoms between October 2016 and February 2017 (0.107, p=0.099), but no correlation with the symptoms in February 2017 (0.073, p=0.289).

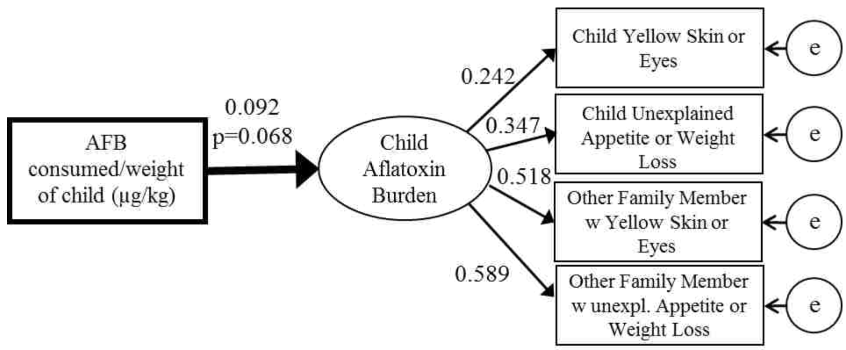


Figure 1. Final basic factor analysis of AFB consumption level on AFB symptom Latent. Arrows are hypothesized causality, rectangles are observed variables, ovals are latent variables, and 'e' are errors. Used DWLS robust estimator; n = 320; Model fit (Chi-square: 8.297, p=0.686; Robust RMSEA: 0.000 (CI: 0.000 – 0.039); Robust CFI: 1.000; Robust TLI: 1.154. Controlled for Child Food Consumption and Socio-economic Status. Additionally, using the 'Composite' of Child AFB symptom and regressing AFB consumption level on it was not statistically significant (Est: 0.142, p=0.364).

3.3. AFB vs EED

The second set of analyses assessed the hypothesized correlation between the AFB consumption level and EED symptoms (composite and latent). For the October 2016 dataset, ordinal regression confirmed a significant correlation between AFB consumption levels and the EED symptom composite at a 5% level (-0.564, p=0.032). Additionally, Figure 3 depicts the factor analysis of the EED symptom latent regressed on by the AFB consumption level. A statistically significant correlation was confirmed between the two factors at the 5% level (-0.093, p=0.036; X2 p=0.122, Robust RMSEA = 0.033 (CI: 0.000 – 0.067); Robust CFI=0.980; Robust TLI=0.960).

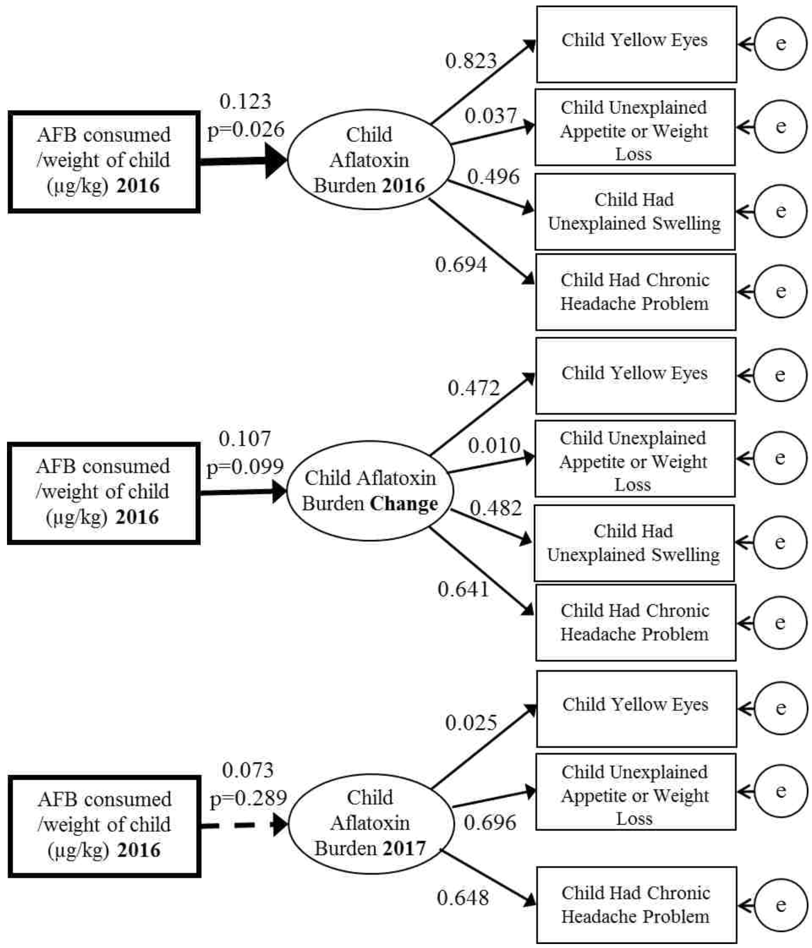


Figure 2. Final factor analysis of AFB consumption level in 2016 on AFB symptom latent for three time points (2016, the difference between 2016 and 2017, and 2017). Displays the change of the correlation between the putative AFB exposure and AFB symptoms change as time between AFB exposure and symptoms increases.

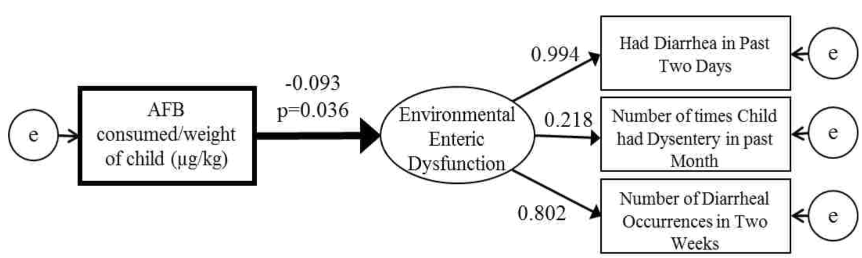


Figure 3. Final factor analysis of AFB consumption level on Environmental Enteric Dysfunction (EED). Used DWLS robust estimator; n = 320; Model fit (Chi-square: 10.067, p=0.122; Robust RMSEA: 0.033 (CI: 0.000 – 0.067); Robust CFI: 0.980; Robust TLI: 0.960. Controlled for Child Food Consumption and Socio-economic Status. Additionally, the ‘Composite’ EED was regressed on by AFB consumption level and was not significant (Est: 0.113, p=0.324).

3.4. AFB Consumption vs EED Symptom Latent vs Height-for-Age

Finally, the hypothesized correlations between the AFB consumption level, the EED symptom latent variable, and the child's height-for-age was analyzed with SEM. Figure 4 depicts the SEM result of the October 2016 dataset with both unstandardized and standardized parameter estimates. First, model fit tests were computed to test the fit of the data to the hypothesized model structure. Tests showed adequate fit (Chi-square: 15.920, p=0.069; Robust RMSEA: 0.049 (CI: 0.000 – 0.092; Robust CFI: 0.945; Robust TLI: 0.889) therefore warranting an analysis of the parameter estimates. The analysis confirmed that the AFB consumption level had a significant correlation with child height-for-age (U: -0.758, S: -0.073, p=0.031) and a significant correlation with the EED symptom latent variable (U: -0.101, S: -0.080, p=0.030). However, the data did not confirm a significant relationship between the EED symptom latent and child height-for-age (U: 0.435, S: 0.053, p=0.429).

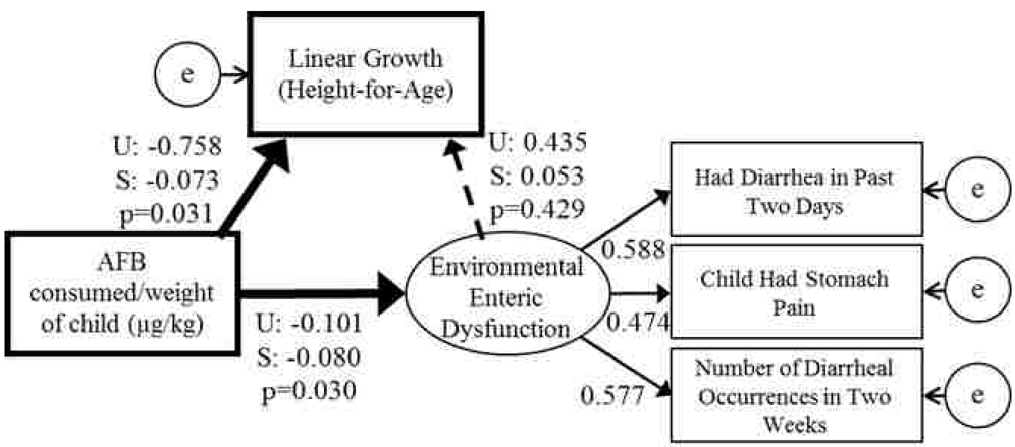


Figure 4. Final structural equation model of AFB consumption level, Environmental Enteric Dysfunction (EED), and child height-for-age. Arrows are hypothesized causality, rectangles are observed variables, ovals are latent variables, and 'e' are errors. Used DWLS robust estimator; n = 320; Model fit (Chi-square: 15.920, p=0.069; Robust RMSEA: 0.049 (CI: 0.000 – 0.092); Robust CFI: 0.945; Robust TLI: 0.889); Controlled for Child Food Consumption and Socio-economic Status.

Utilizing the two time-point data set, the Kruskal-Wallis tests confirmed that the dichotomous response of the child putatively consuming over 10 ng/kg body weight of AFB in October 2016 had a significant correlation with child height-for-age in October

2016 (p=0.080) and the change between the child height-for-age in October 2016 and February 2017 (p=0.074), but not in February 2017 (p=0.763).

Finally, Figure 5 depicts the final SEM of the hypothesized correlations between AFB consumption levels, EED in October 2016, EED in February 2017, and the change in child height-for-age between October 2016 and February 2017 (catch-up growth). Tests of model fit showed adequate fit permitting the assessment of the parameter estimates (Chi-square: 22.666, p=0.750; Robust RMSEA: 0.000 (CI: 0.000 – 0.046); Robust CFI: 1.000; Robust TLI: 1.160). The AFB consumption level had two confirmed significant correlations; first with catch-up growth (S: -2.359, U: -0.161, p=0.009) and second with EED in February 2017 (S: -0.216, U: -0.129, p=0.084). There was no confirmed significant correlation between the EED symptoms nor between either EED symptoms and child height-for-age.

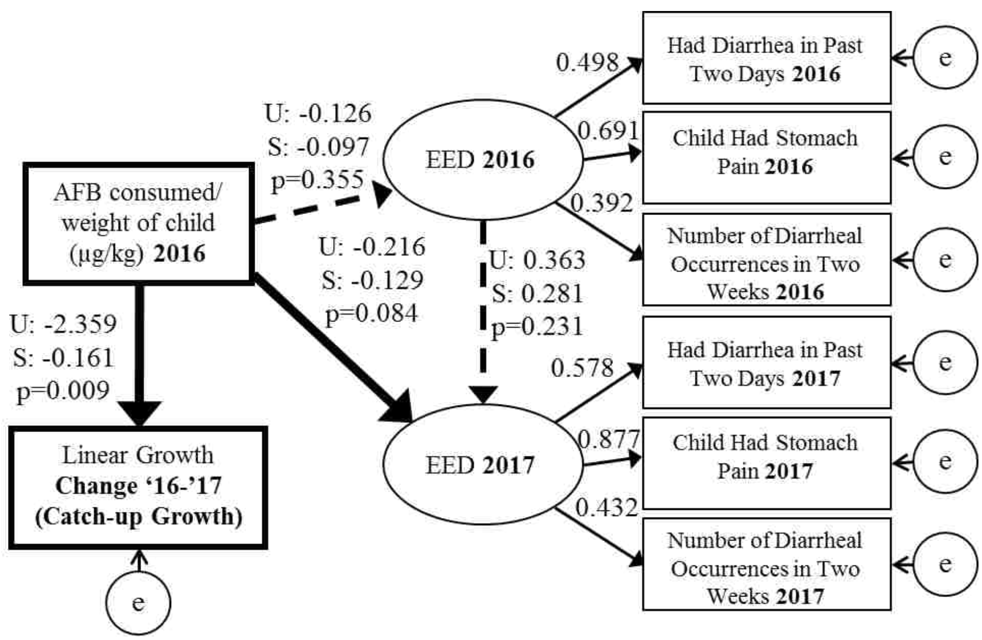


Figure 5. Final two time point structural equation model of AFB consumption level, Environmental Enteric Dysfunction (EED) in 2016 and 2017, and the change in child height-for-age between 2016 and 2017. Arrows are hypothesized causality, rectangles are observed variables, ovals are latent variables, and 'e' are errors. Used DWLS robust estimator; n = 120; Chi-square: 22.666, p=0.750; Robust RMSEA: 0.000 (CI: 0.000 – 0.046); Robust CFI: 1.000; Robust TLI: 1.160. Controlled for Child Food Consumption and Socio-economic Status.

4. DISCUSSION

4.1. AFB and Symptoms

This study explored alternative methods for imputing exposure levels of AFB on children in Guatemala. The AFB consumption level value attempted to capture exposure through a computed consumption level. This method is used in diet diversity and micronutrient studies in that one assumes the reported single-day level of consumption of the subject represents the mean level of consumption when analyzed among a population.[17] Doak et al utilized a similar method when analyzing calorie and nutrient intake among children in Guatemala.[26] The final AFB symptom variable value was based on the most frequent symptoms previously associated with chronic levels and high levels of AFB exposure.[19] Bosa et al discuss the potential symptoms of AFB including jaundice and appetite loss.[18]

To validate an AFB symptom approach in assessing AFB exposure levels, this study analyzed the correlation between the AFB consumption level (input) and the AFB symptom variable (output). The data showed a significant correlation between exposure (a continuous variable) and symptoms (yes-no responses) for the October 2016 dataset. Several animal studies have demonstrated similar correlations between exposure levels and symptoms, Williams et al review studies that demonstrated correlations between increased AFB exposure among mice and chickens and liver-based dysfunctions.[7] Within the human population Jolly et al found a correlation between AFB exposure and vomiting and abdominal swelling.[19] Mapesa et al reviews results from a variety of studies on AFB to build a hypothetical causal diagram of mechanistic pathways leading to symptoms.[5] Additionally, the study analyzed this correlation as the period between putative AFB consumption and AFB symptoms increased (change between 2016/2017 and symptoms in 2017). The data suggest the validity of the correlation becomes weaker as time increases. Hinton et al reported that in rats dosing of AFB was correlated with a closely followed peak of immune stimulation suggesting short periods of separated exposure and effect.[27]

The limitations with these two indicators include, as mentioned, numerous mechanistic intestinal changes along the causal pathway that need further investigation. Additionally, symptoms used for AFB exposure were associated with liver dysfunction and not intestinal exposure. Potential confounding factors include birth related jaundice, limited understanding among mothers of AFB symptom diagnosis, and other related factors to appetite or weight loss. Finally, while this supports recent findings, spurious correlations are possible among the data and therefore this correlation needs further validation for significance.

4.2. AFB and Health

The analysis confirmed significant correlations between AFB consumption levels and EED symptoms and was tested using multiple statistical techniques. However, the results demonstrated a consistent negative correlation meaning as AFB consumption levels increased EED problems decreased. This was counter to the original hypothesis. Both the AFB consumption level and the AFB symptoms (latent and composite) were negatively correlated with the EED symptoms. Previous findings utilizing animal models had reported alterations in intestinal functionality similar to EED.[10] Applegate et al and Yunus et al found evidence of reduced intestinal absorptive capacity in chickens when exposed to high levels of AFB.[28,29] However, limited human models have been tested to date. Smith et al has published a conceptual framework for the effect of AFB on child height-for-age including intestinal disruption.[30] The primary results from the AFB animal exposure studies reported an increase in immune response activity (overstimulation of cytokines).[10] Furthermore, the EED symptoms potentially captured more severe cases of EED, as the occurrence of diarrhea and multiple bouts of diarrhea may be associated with more severe cases of poor intestinal integrity as compared to symptoms such as stomach pain.

The hypothesized correlations between AFB, EED, and child height-for-age were assessed using SEM for both datasets. The analysis of the October 2016 dataset confirmed a significant correlation between AFB and child height-for-age; however, no significant correlation was confirmed between EED and child height-for-age. Again, there was a confirmed negative correlation between AFB and EED. Similar results were identified when the two time-point dataset was analyzed. AFB, again, was correlated negatively to both child height-for-age (confirming hypothesis) and EED in February 2017 (counter to hypothesis). Previous research suggests a relationship between EED and child height-for-age.[3] However, the interaction between AFB, EED, and child height-for-age is less clear. This study supports the hypothesis that AFB is correlated with child height-for-age; however, the specific mechanistic pathways of this relationship and the functionality of the intestinal tract of children is less well known. Further research should be undertaken attempting to validate the AFB-child height-for-age mechanistic pathway and elucidate the AFB-EED pathway. Limitations for this portion of the study included accurate reporting of occurrences of diarrhea, dysentery, stomach pain, and other EED related symptoms.

Finally, the results from the analyses of the hypothesized correlations that included either the AFB or EED symptoms supported the use of the latent variable mathematical theory used in factor analyses and SEM. A composite score assumes 1) the ‘indicator’ variables used to create the composite (e.g. for AFB; yellow eyes, loss of appetite, etc.) explain all of the composite factor in its entirety and 2) causality runs from the indicator variables to the composite score. In latent variable theory 1) it is possible to switch out indicator variables and maintain the integrity of the latent variable and 2) causality runs from the latent variable to the indicator variables.[31] Symptoms, whether from AFB or EED, are caused by an underlying dysfunction or problem, which is supported by the latent variable theory of causality and was supported in this study based on the consistent improvements in identifying potential correlations among variables.

This study supports the hypothesis of correlations between 1) AFB consumption levels among children and the potential AFB related symptoms and 2) AFB consumption levels among children and the child height-for-age. With emerging concerns around both EED’s and AFB’s role in child development, it is critical to understand how each one affects child growth. While further research will be needed to investigate specific mechanistic pathways between EED and AFB, practitioners in AFB prone countries must be aware of the potential harmful effects on child health.

REFERENCES

1. Black RE, Victora CG, Walker SP, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. Lancet. 2013;382(9890):427-451. doi:10.1016/S0140-6736(13)60937-X.

2. United Nations General Assembly. Transforming Our World: The 2030 Agenda for Sustainable Development. Vol A/70/L.1.; 2015:1-35.

3. Keusch GT, Rosenberg IH, Denno DM, et al. Implications of acquired environmental enteric dysfunction for growth and stunting in infants and children living in low-and middle-income countries. Food Nutr Bull. 2013;34(3):357-364.

4. Ahmed T, Auble D, Berkley J a, et al. An evolving perspective about the origins of childhood undernutrition and nutritional interventions that includes the gut microbiome. Ann N Y Acad Sci. 2014;1332:22-38. doi:10.1111/nyas.12487.

5. Mapesa JO, Maxwell AL, Ryan EP. An Exposome Perspective on Environmental Enteric Dysfunction. Environ Health Perspect. 2016;124(8):1121-1126.

6. Smith LE, Stoltzfus RJ, Prendergast a. Food Chain Mycotoxin Exposure, Gut Health, and Impaired Growth: A Conceptual Framework. Adv Nutr An Int Rev J. 2012;3(4):526-531. doi:10.3945/an.112.002188.

7. Williams JH, Phillips TD, Jolly PE, Stiles JK, Jolly CM, Aggarwal D. Human aflatoxicosis in developing countries: a review of toxicology, exposure, potential health consequences, and interventions. Am Soc Clin Nutr. 2004;80(June 2015):1106-1122.

8. International Agency for Research on Cancer. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Aflatoxins.; 2006.

9. Gourama H, Bullerman LB. Aspergillus flavus and Aspergillus parasiticus, aflatoxigenic fungi of concern in foods and feed – a review. J Food Prot. 1995;58:1395-1404.

10. Wild CP, Miller JD, Groopman JD. Mycotoxin Control in Low- and Middle-Income Countries.; 2016.

11. LIzarraga-Paulin EG, Moreno-Martinez E, Miranda-Castro SP. Chapter 13: Aflatoxins and their Impact on Human and Animal Health: An Emerging Problem. In: Aflatoxins - Biochemistry and Molecular Biology. ; 2010:1-28.

12. U.S. Food and Drug Administration. Food and Drug Administration Compliance Program Guidance Manual: Mycotoxins in Domestic and Imported Foods.; 2016:1-27. doi:10.1300/J138v04n01_14.

13. European Commission. Commission Regulation (EC) No 1881/2006 of 19 December 2006 Setting Maximum Levels for Certain Contaminants in Foodstuffs.; 2006:1-35. doi:10.2203/dose-response.06-012.Hanekamp.

14. Torres O, Matute J, Gelineau-van Waes J, et al. Human health implications from co-exposure to aflatoxins and fumonisins in maize-based food in Latin America: Guatmeala as a Case Study. World Mycotoxin J. 2015;8(2):143-159.

15. United States Agency International Development. Baseline Study for the Title II Development Food Assistance Programs in Guatemala.; 2014.

16. Torres O, Matute J, Gelineau-van Waes J, et al. Human health implications from co-exposure to aflatoxins and fumonisins in maize-based foods in Latin America: Guatemala as a case study. World Mycotoxin J. 2015;8(2):143-159. doi:10.3920/WMJ2014.1736.

17. Swindale A, Bilinsky P. Household Dietary Diversity Score for Measurement of Household Food Access: Indicator Guide. Washington, DC; 2006. http://www.fantaproject.org/sites/default/files/resources/FAFSA_April2003.pdf.

18. Bbosa GS, Kitya D, Lubega A, Ogwal-Okeng J, Anokbonggo WW, Kyegombe DB. Chapter 12: Review of the Biological and Health Effects of Aflatoxins on Body Organs and Body Systems. In: Aflatoxins - Recent Advances and Future Prospects. ; 2013:239-265.

19. Jolly PE, Jiang Y, Ellis WO, et al. Association between aflatoxin exposure and health characteristics, liver function, hepatitis and malaria infections in Ghanaians. J Nutr Environ Med. 2007;16(3-4):242-257. doi:10.1080/13590840701703918.

20. Korpe PS, Petri W a. Environmental enteropathy: Critical implications of a poorly understood condition. Trends Mol Med. 2012;18(6):328-336. doi:10.1016/j.molmed.2012.04.007.

21. Gong YY, Egal S, Hounsa A, et al. Determinants of aflatoxin exposure in young children from Benin and Togo, West Africa: The critical role of weaning. Int J Epidemiol. 2003;32(4):556-562. doi:10.1093/ije/dyg109.

22. McKight PE, Najab J. Kruskal-Wallis Test. Corsini Encycl Psychol. 2010;1.

23. Armstrong BG, Sloan M. Ordinal regression models for epidemiologic data. Am J Epidemiol. 1989;129(1):191-204. doi:10.1093/aje/kwp107.

24. Mîndrilã D. Maximum likelihood (ML) and diagonally wighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. Int J Digit Soc. 2010;1(1):60-66. doi:10.20533/ijds.2040.2570.2010.0010.

25. Grace JB. Structural Equation Modeling and Natural Systems. First. United Kingdom: Cambridge University Press; 2006.

26. Doak CM, van der Starre RE, van Beusekom I, Campos Ponce M, Vossenaar M, Solomons NW. Earlier introduction of aguitas is associated with higher risk of stunting in infants and toddlers in the Western Highlands of Guatemala. Am J Clin Nutr. 2013;97(3):631-636. doi:10.3945/ajcn.112.047621.INTRODUCTION.

27. Hinton DM, Myers MJ, Raybourne RA, et al. Immunotoxicity of aflatoxin B1 in rats: Effects on lymphocytes and the inflammatory response in a chronic intermittent dosing study. Toxicol Sci. 2003;73(2):362-377. doi:10.1093/toxsci/kfg074.

28. Applegate TJ, Schatzmayr G, Prickel K, Pricket K, Troche C, Jiang Z. Effect of aflatoxin culture on intestinal function and nutrient loss in laying hens. Poult Sci. 2009;88(6):1235-1241. doi:10.3382/ps.2008-00494.

29. Yunus a W, Ghareeb K, Abd-El-Fattah a a M, Twaruzek M, Böhm J. Gross intestinal adaptations in relation to broiler performance during chronic aflatoxin exposure. Poult Sci. 2011;90(8):1683-1689. doi:10.3382/ps.2011-01448.

30. Smith LE, Stoltzfus RJ, Prendergast A. Food chain mycotoxin exposure, gut health, and impaired growth: A conceptual framework. Adv Nutr. 2012;3(4):526-531. doi:10.3945/an.112.002188.

31. Kline RB. Principles and Practice of Structural Equation Modeling. 3rd ed. New York: The Guilford Press; 2011.

IV. MAIZE STORAGE, POST-HARVEST PRACTICES, AND MARKET PURCHASE HABITS ARE CORRELATED WITH REPORTED SYMPTOMS OF AFLATOXIN EXPOSURE AMONG CHILDREN IN SAN VICENTE, GUATEMALA

To be submitted to the Journal of Environmental Health Perspectives

Lee E. Voth-Gaeddert1, Matthew Stoker2, Daniel B. Oerther1

1 Civil, Architectural, and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO, USA, lv6w3@mst.edu

2 Independent Researcher

ABSTRACT

Background: The fungus Aspergillus produces aflatoxins that are classified as a group 1 carcinogen by the World Health Organization (WHO). Prior research documented elevated levels of aflatoxins in maize samples from markets within 12 of 24 departments throughout Guatemala. In the current study, cross-sectional data collected in October 2016 and February 2017 were used to test hypothesized correlations within two models that incorporated variables hypothesized to contribute to increased exposure to aflatoxins from maize purchased from local markets or from subsistence maize production.

Methods: Health assemblies were held by local health practitioners for mothers with children between one month and five years of age in October 2016, one month before harvest, and February 2017, two months after harvest. At the assemblies, surveys were administered orally to mothers in the local dialect by translators. Immediately following, house visits were conducted with mothers who attended the health assemblies to collect samples of maize allocated for consumption. The level of aflatoxin in the maize sample was determined using an enzyme-linked immunosorbent assay (ELISA) method. For October 2016, an odds ratio and relative risk value of having maize with aflatoxin levels greater than 15 parts per billion were determined for households who purchased maize from local markets as opposed to households with subsistence maize production. Structural equation modeling (SEM) was then used to analyze two hypothesized models for October 2016 and two for February 2017 exploring the putative routes of exposure from either maize purchased from local markets or from subsistence maize production.

Findings: The results confirmed the hypothesis that households that purchased maize from the market had 3.31 higher odds (95% CI: 1.35-8.11) and 2.16 times the likelihood (95% CI: 0.98-4.71) of having maize with levels of aflatoxin above 15 ppb in their house compared to subsistence households. The October 2016 SEM for market-purchased maize confirmed that purchase habits had a negative significant effect size on the child aflatoxin burden (-0.220, p=0.037). The October 2016 SEM for subsistence maize confirmed that post-harvest practices, observing fungus, and the type of maize storage had significant negative effect sizes on child aflatoxin burden (-0.158, p=0.048; -0.111, p=0.004; and -0.082, p=0.024 respectively). The February 2017 SEMs for market-purchased maize and subsistence maize confirmed observing fungus (-0.391, p=0.000) and higher maize price for longer storage (0.079, p=0.089) were significant, respectively. Additionally, at both time points households who reported receiving a higher maize price for longer storage also reported having improved storage facilities (2016: 0.063, p=0.001 and 2017: 0.230, p=0.017).

Interpretation: This is the first study to report on correlations between a set of variables associated with the potential transmission of maize-born aflatoxins specific to Guatemala. Based on the results multiple interventions may be effective, but varying in effectiveness depending on the time of year and sources of maize for individual households. To reduce aflatoxin exposure to children, practitioners and policy makers should consider all options including market-based and educational interventions.

Keywords: Aflatoxins, Purchase Habits, Maize Storage, Guatemala, Child Health, Structural Equation Modeling

1. INTRODUCTION

Members of the fungal genus, Aspergillus spp., biochemically produce aflatoxins that are classified as a group 1 carcinogen by the World Health Organization (WHO) (International Agency for Research on Cancer, 2006). Aflatoxins are the most prevalent and the most toxic type of mycotoxin and are found to grow on crops including maize, sorghum, cassava, and ground nuts (LIzarraga-Paulin, Moreno-Martinez, & Miranda-Castro, 2010). Among the peoples of Guatemala, maize, in the form of tortillas, tamales, and soup, is a staple of the diet (United States Agency International Development, 2014). Recent estimates suggest that 72% of the daily energy intake from food among the people of Guatemala come from Maize (Agriculture and Consumer Protection, n.d.). Additionally, prior studies have documented the wide spread occurrence of elevated levels of aflatoxins in the maize supplies in both public markets as well as private households throughout Guatemala (Torres et al., 2015). This combination of factors, namely, the widespread consumption of maize and the widespread contamination of maize with aflatoxin, may create a significant public health threat to the people of Guatemala.

The International Agency for Research on Cancer (IARC) recently released a meta study documenting the results of research describing the effects of aflatoxins on child linear growth arguing for an increase in research to determine the potential routes of environmental exposure (Wild, Miller, & Groopman, 2016). Guatemala has the sixth highest child stunting rate in the world and the worst rate of child stunting in the western hemisphere at 49% of all children under five years of age stunted (United Nations Children’s Emergency Fund, 2013). Stunting is defined as a height-for-age score of at least two standard deviations below the WHO growth mean (United Nations General Assembly, 2015). Despite significant efforts to address the issue of child stunting in Guatemala, the condition persists.

To reduce the potential for aflatoxin exposure to children, identifying the key mechanisms by which the fungus Aspergillus becomes prevalent in maize is important. Fungal growth may occur in the field, during harvest, during post-harvest practices, in storage, and in transport (Wild et al., 2016). The majority of households in the western highlands of Guatemala rely on subsistence maize but may supplement their stocks with purchases from local markets during the lean season (United States Agency International Development, 2012). The two primary options households have for obtaining maize include subsistence farming or purchasing maize from the market. Each scenario has unique potential fungal toxin growth and transmission pathways. To identify potential intervention points for the reduction of aflatoxin exposure, modeling the systems at a household level can provide useful insight.

In this study, an analysis was conducted on data collected from San Vicente, Guatemala in October 2016 and February 2017 to assess hypothesized correlations of influences in the transmission of aflatoxins using two structural equation models at each time point. The two hypothesized models included; 1) factors related with the local market that may influence children to display aflatoxin exposure symptoms and 2) factors related with household maize production that may influence children to display aflatoxin exposure symptoms. Results confirmed that reported aflatoxin exposure symptoms were correlated with the type of maize storage and post-harvest practices for subsistence households while improved market purchase habits were significant for households acquiring maize from the market.

2. METHODOLOGY

2.1. Location

San Vicente, Guatemala (15 1’33.20N, 91 35’1.99W) is located in the western highlands at an elevation of 2,780 meters, with an average range of temperatures of 5.1C to 17.0F, and an annual rainfall of 1,310 mm. The farming of maize is the primary source of income for the majority of households and includes only one harvest season which occurs in November. The dominate cultural identity within the region is Mayan with the primary language being Quiché and the secondary language being Spanish.

2.2. Data Collection and Preparation

In October of 2016 and February of 2017, health professionals from the Health Center of San Vicente conducted health assemblies for mothers with children under five years of age. Households were informed of the health assemblies via community wide public announcements and flyers. During each assembly, surveys were administered orally to the mothers in their local dialect via a translator. Subsequently, after each health assembly, house visits were conducted with participants of the assemblies to collect household observations and samples of maize from the household’s storage allocated for consumption. Institutional Review Board approval was attained from Missouri University of Science and Technology to analyze the de-identified data collected by the local health center. The names and descriptions of the variables that were included in the survey are shown in Table 1.

Table 1. Variables and descriptions used in the structural equation models for the market and subsistence models.

Variables	Definitions
Child Yellow Eyes	The mother reported the child having yellow coloring in the eyes
Child Unexplained Appetite or Weight Loss	The mother reported the child losing weight or appetite unexplainably
Child Headache	The mother reported the child having problems with chronic headaches
Child Had Unexplained Swelling	The mother reported the child having unexplained swelling in any extremities of the body
Observed Fungus in Maize	Field worker observed fungus in household maize storage
Market Model	
Market Purchase Habits	Score based on purchase habits most associated with potential fungal growth
Remoteness of Maize Market	Transport time from Quetzaltenango to given market
Subsistence Model	
Type of Maize Storage	Observed type of storage used by household for maize
Amount of Time Used for Drying Maize	Reported time used for drying maize during previous harvest
Type of Drying Surface for Maize	Reported type of surface used in the drying of maize
Type of Practice Used to Shear Maize Form Cob	Reported type of method used to remove kernels from cob
Receive More $$ for Better Quality Maize	House reports receiving more money for selling better quality maize
Higher Value of Maize if Store Longer	House reports receiving more money for delaying the selling of the maize after harvest

Samples of maize collected from households were sealed in paper bags and immediately sent to Guatemala City to determine alfatoxic levels using a commercially available enzyme-linked immunosorbent assay (ELISA) according to the manufacturer's instructions. The test identified the parts per billion of aflatoxin present in the sample. The protocol reported by the Neogen Corporation was utilized (Neogen Corporation, 2012). Additionally, during maize sampling the household was asked specifically about the origin of the maize sampled, which was recorded and utilized for computing a relative risk and odds ratio.

From the two data collection campaigns (October 2016 and February 2017) four datasets were created to be analyzed by the four SEMs. For each time point, two subsamples were drawn from all represented households at the health assemblies based on responses to subsistence and market attendance questions in the survey. The subsistence subsample was selected based on if the majority of maize consumed in the past month was obtained from subsistence farming. The market attendance subsample was selected based on if the household had acquired maize at any point in the previous month from a market. This meant that the same household could be included in both datasets; however, if both maize sources were utilized, both sources may have contributed to child aflatoxin exposure therefore warranting this method of data subsampling.

2.3. Statistical Approaches

The levels of aflatoxin measured in the samples of maize collected from the households were used to calculate the relative risk and odds of a household having a high level of aflatoxin in their maize storage based on the specific source of acquisition (market versus subsistence). The aflatoxin limit denoted as ‘high level’, was set at 15 ppb based on the United States and European Union import regulation levels of 20 parts per billion and 5 parts per billion (European Commission, 2006; U.S. Food and Drug Administration, 2016). From the measured aflatoxin level of the sample and the recorded responses of the specific origin of the maize sampled, a relative risk and odds could be computed. Specific discussion on the methodology for calculating relative risk and the odds ratio can be found in Daniel 1995 (Daniel, 1995).

The second statistical approach used in this study was structural equation modeling (SEM). SEM is a statistical technique that combines path analysis and factor analysis to analyze multiple interacting hypotheses, simultaneously. Factor analysis is used to compute latent variables from a set of hypothesized indicator (manifest) variables. Figure 1 depicts the two hypothesized SEMs for this study. In the subsistence SEM two latent variables, denoted as ovals labeled Child Aflatoxin Burden and Post-Harvest Practices, are included (just Child Aflatoxin Burden in market SEM). The single headed arrows reflect hypothesized causality and the rectangles denote directly observable variables. Path analysis utilizes a covariance matrix approach to compare the fit of the data (all observable and latent variables) to the fit of the hypothesized model.

Four model fit statistics are used to measure adequate fit. These include Chi-Square (X2, p>0.05), Root Mean Square Error of Approximation (RMSEA<0.08) Confirmatory Fit Index (CFI>0.90), and the Tucker-Lewis Fit Index (TLI>0.90). Due to the ordinal and dichotomous nature of the variables the estimator diagonally weighted least squares was utilized (Mîndrilã, 2010). If adequate fit is attained, parameter estimates are then analyzed and are given in standardized (S) and unstandardized (U) regression formats. Figure 1 displays the two models reflecting putative child aflatoxin exposure from maize purchased from local markets and from subsistence maize production. Further reading for SEM can be found in Grace 2006 (Grace, 2006).

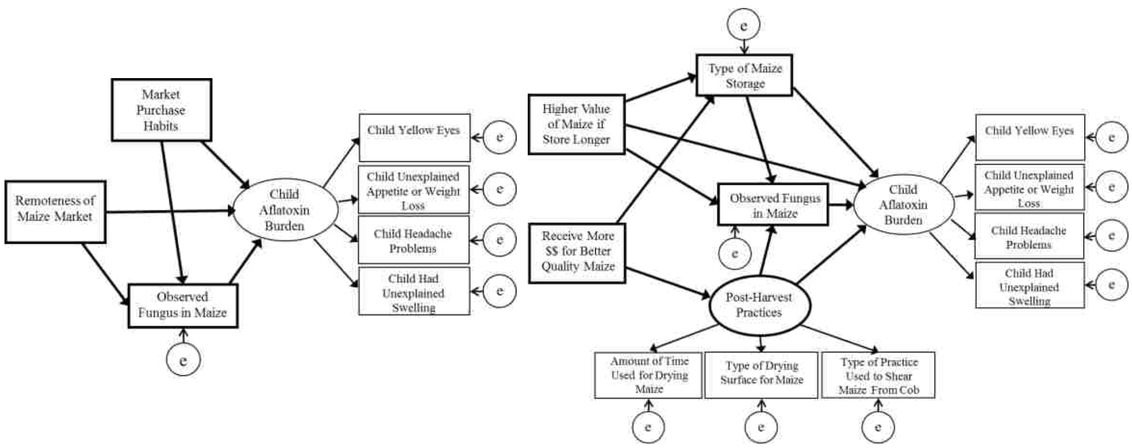


Figure 1. Hypothesized structural equation models for AFB exposure routes from the Market Model (on the left) and Subsistence Model (on the right). Arrows are hypothesized causalities, rectangles are observable variables, ovals are latent variables, and 'e' are errors.

3. RESULTS

3.1. Descriptive Statistics

Table 2 displays the descriptive statistics for the data collected in San Vicente in October 2016 and February 2017, divided into subsistence households and market households for each time point. First, the average level of tested aflatoxin in the maize samples from October 2016 was 7.74 ppb (range: 0-96ppb, n=229) with 9.6% of households having 15 ppb or greater levels of aflatoxin. Of these high-level households (9.6%) only 9.1% (n=2) had observed fungus in their maize storage. For the SEM datasets 50% of households surveyed in February 2017 had not attended the health assemblies in October 2016. Overall, in October 2016, 45% of households reported acquiring some maize from the market within the past year while 55% of households reported this in February 2017. Datasets included subsistence (n=281) and market attending (n=174) households in October 2016 and subsistence (n=160) and market attending (n=168) households in February 2017. In October 2016, 13.5% of households were included in both datasets, while 8.5% of households were included in both datasets for February 2017. Lastly, fungus was observed in the maize of 8.3% of households in October 2016 and 13.7% in February 2017.

Table 2. Descriptive statistics for the October 2016 and February 2017 datasets.

Variable	October 2016 Dataset	February 2017 Dataset
Subsistence Sample size	281 children	160 children
Market Sample Size	174 children	168 children
AFB in maize samples*	7.74 ppb (0-96ppb)	-
Reported acquiring maize from market in past month	45%	55%
Children in both datasets	13.5%	8.5%
Observed fungus in maize in household	8.3%	13.7%
*subsample of total; n=229		

3.2. Odds Ratio and Relative Risk

The results of the odds ratio and relative risk of having 15 ppb of aflatoxin or greater in the sample of household maize is shown in Table 3. For October 2016 (n=229), households that reported acquiring the maize sample from the market had 3.31 (95% CI: 1.35-8.11) higher odds or were 2.16 (95% CI: 0.98-4.71) times more likely to have 15 ppb or great aflatoxin levels in their maize sample then subsistence households.

Table 3. The odds ratio and relative risk ratio of a household who attended the market having a maize sample with 15 ppb of aflatoxin or higher as identified by an ELISA test; n=229.

	Market	Subsistence	Total
AFB≥15ppb	11	48	59
AFB<15ppb	11	159	170
	22	207	229
	0.50	0.23	
Relative Risk	2.16	**Odds Ratio**	3.31

3.3. Market Maize

The final SEM associated with putative child aflatoxin exposure from the market maize in October 2016 is depicted in Figure 2. The data confirmed adequate fit to the hypothesized model based on the model fit statistics (Chi-square: 13.439, p=0.266; RMSEA: 0.036 (CI: 0.000 – 0.100); Robust CFI: 0.944; Robust TLI: 0.898) warranting the investigation of the parameter estimates.

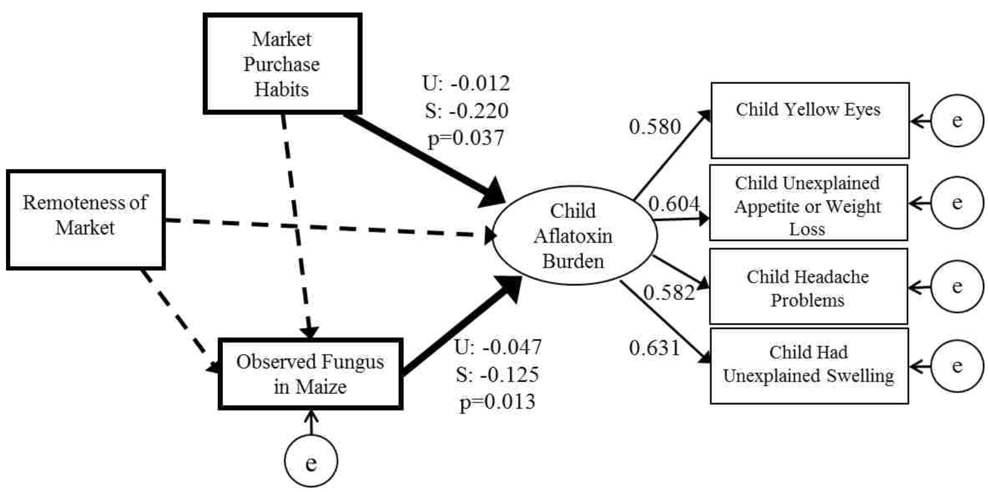


Figure 2. Final structural equation model of October 2016 market model. DWLS robust estimator used; n = 174, Chi-square: 13.439, p=0.266; RMSEA: 0.036 (CI: 0.000 – 0.100); Robust CFI: 0.944; Robust TLI: 0.898. Arrows are hypothesized direction, rectangles are observed variables, ovals are latent variables, 'e' are error. Solid arrows are confirmed statistically significant correlations at a 10%, dashed arrows are correlations important to the overall SEM but not significant at a 10% level. Size added for emphasis. S = standardized parameter estimate, U = unstandardized parameter estimate, p = statistically significant level.

The latent variable, Child Aflatoxin Burden, was regressed on three indicator variables that were hypothesized to be outcomes (i.e. symptoms) of the underlying problem (i.e. high aflatoxin exposure). Indicators included the ‘yes/no’ responses of four symptom-based questions related to aflatoxin exposure including, has the child had yellow eyes, has the child had unexplained appetite or weight loss, has the child had problems with headaches, and has the child experienced unexplained swelling (Bbosa et al., 2013; Voth-Gaeddert, Stoker, Torres, & Oerther, n.d.; Wild et al., 2016). Two variables were correlated with the Child Aflatoxin Burden latent variable including the purchase habits of the mother for maize at the market (Purchase Habits) and the observed presence of fungus in the household maize sample (Observed Fungus). Purchase Habits was significant at a 1% level with a standardized effect size of -0.220 (p=0.037) while Observed Fungus was also significant at a 1% level with a standardized effect size of -0.125 (p=0.013). Additionally, remoteness of the market (Market Remoteness) was not statistically significant with either Child Aflatoxin Burden or Observed Fungus, but contributed to the overall fit of the model to the data.

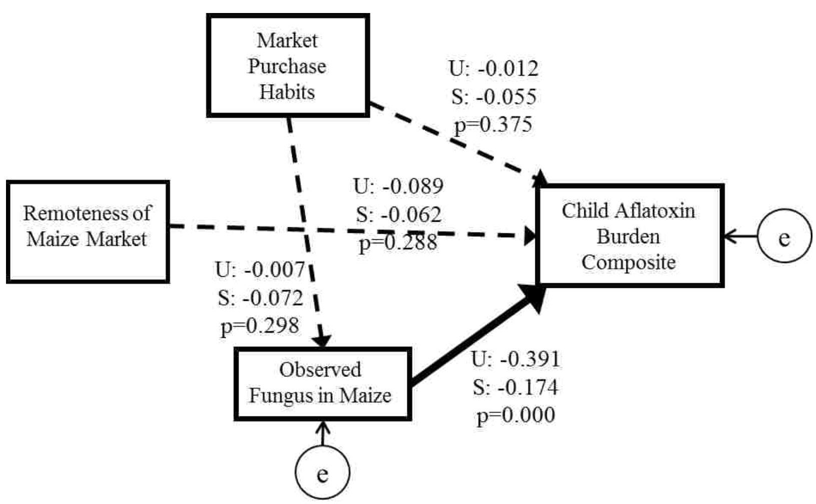


Figure 3. Final structural equation model of February 2017 market model. DWLS robust estimator used; n=168; Chi-square: 0.182, p=0.670; RMSEA: 0.000 (CI: 0.000 – 0.000); Robust CFI: 1.000; Robust TLI: 1.218.

The final SEM associated with putative child aflatoxin exposure from the market in February 2017 is depicted in Figure 3. The structure of the Child Aflatoxin Burden latent variable was unstable among the specific population the data was drawn from and therefore warranted the use of the composite form of the variable so the model could be estimated. For comparison, the Child Aflatoxin Burden variable from the October 2016 model was changed to a composite variable and the model reassessed. Results showed this change did not affect the order of effect sizes among variables and the specific effect sizes did not vary significantly (-Δ 0.094, -Δ p=0.009). For the final market SEM of February 2017, utilizing the composite variable (Figure 4), the model fit statistics were adequate (Chi-square: 0.182, p=0.670; RMSEA: 0.000 (CI: 0.000 – 0.000); Robust CFI: 1.000; Robust TLI: 1.218) warranting the assessment of the parameter estimates. Only one correlation was confirmed as significant within the model, Observed Fungus, which was significant at a <0.01% level and had a negative standardized parameter estimate of -0.174 (p=0.000).

3.4. Subsistence Maize

Figure 4 depicts the final SEM associated with putative child aflatoxin exposure from a household's subsistence maize in October 2016. In addition to the Child Aflatoxin Burden latent variable, a Post-Harvest Practices latent variable was hypothesized. The indicator variables for this latent included the amount of time the maize was dried (often in the sun), the type of surface used for drying, and the practice used for removing the maize kernels from the cob. The model fit statistics confirmed good fit of the data to the model (chi square: 34.786, p=0.144; Robust RMSEA: 0.028 (CI: 0.000 – 0.053); Robust CFI: 0.919; Robust TLI: 0.869) permitting the analysis of the parameter estimates.

Variables confirmed as significantly correlated with the Child Aflatoxin Burden latent variable included Post Harvest Practices, Observed Fungus, the type of household maize storage (Maize Storage) and higher maize price for longer storage (Storage Profit). Post-Harvest Practices was significant at a 5% level with a standardized effect size of -0.158 (p=0.048). Observed Fungus was significant at a 0.5% level with a standardized effect size of -0.111 (p=0.004) while Maize Storage was significant at a 5% level with an effect size of -0.082 (p=0.024). Finally, Storage Profit was significant at a 5% level with a standardized effect size of 0.068 (p=0.040).

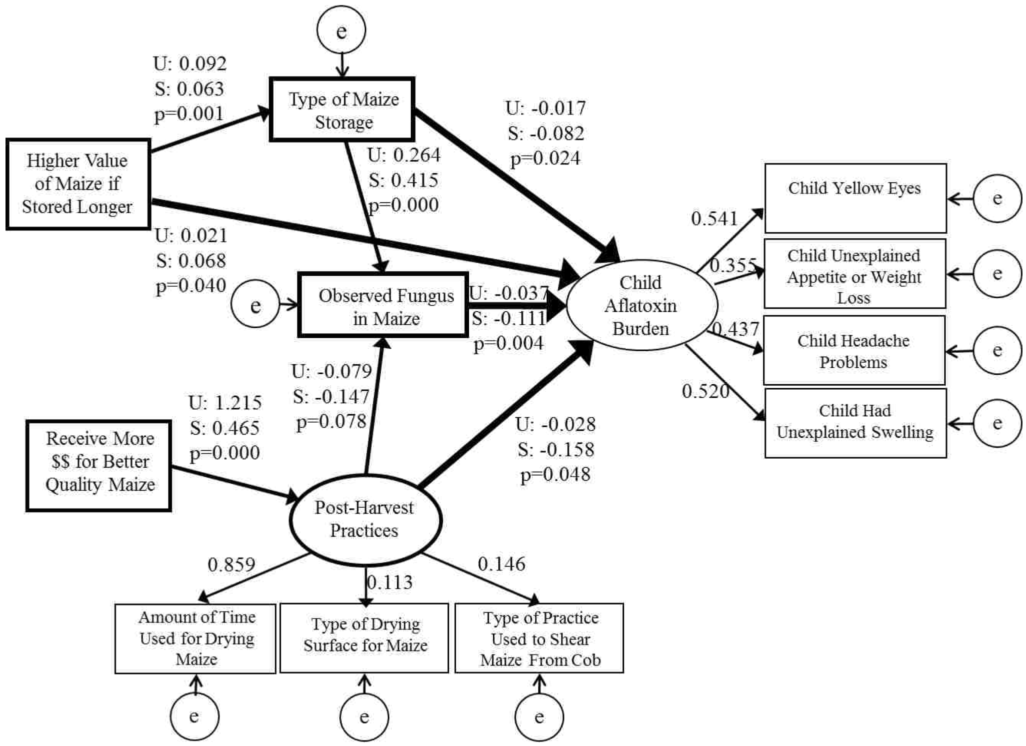


Figure 4. Final structural equation model of October 2016 subsistence model. DWLS robust estimator used; n=281; Chi-square: 34.786, p=0.144; Robust RMSEA: 0.028 (CI: 0.000 – 0.053); Robust CFI: 0.919; Robust TLI: 0.869.

Additionally, Maize Storage and Post-Harvest Practices had indirect effects on Child Aflatoxin Burden through the mediation of Observed Fungus. Maize Storage had a significant correlation with Observed Fungus at a 0.1% level with a standardized effect size of 0.415 (p=0.000). Post-Harvest Practices had a significant correlation with Observed Fungus at a 10% level with a standardized effect size of -0.147 (p=0.078). Finally, two hypothesized correlations that included two market variables were identified as significant. Storage Profit was correlated with Maize Storage at a 0.01% level with a standardized effect size of 0.063 (p=0.001). Receiving a higher value for maize based on the quality (Quality Profit) was correlated with Post-Harvest Practices at a <0.01% level with a standardized effect size of 0.465 (p=0.000).

The final SEM associated with putative child aflatoxin exposure from a household's subsistence maize from February 2017 is depicted in Figure 5. The results from the model fit statistics confirmed good fit between the hypothesized model and the data (Chi-square: 51.497, p=0.057; Robust RMSEA: 0.039 (CI: 0.000 – 0.063); Robust

CFI: 0.773; Robust TLI: 0.668) permitting the analysis of the parameter estimates. The only variable correlated with Child Aflatoxin Burden was Storage Profit at a 10% level with a standardized parameter estimate of 0.079 (p=0.089). Storage Profit also had a significant relationship with Maize Storage at a 5% level with a standardized parameter estimate of 0.230 (p=0.017).

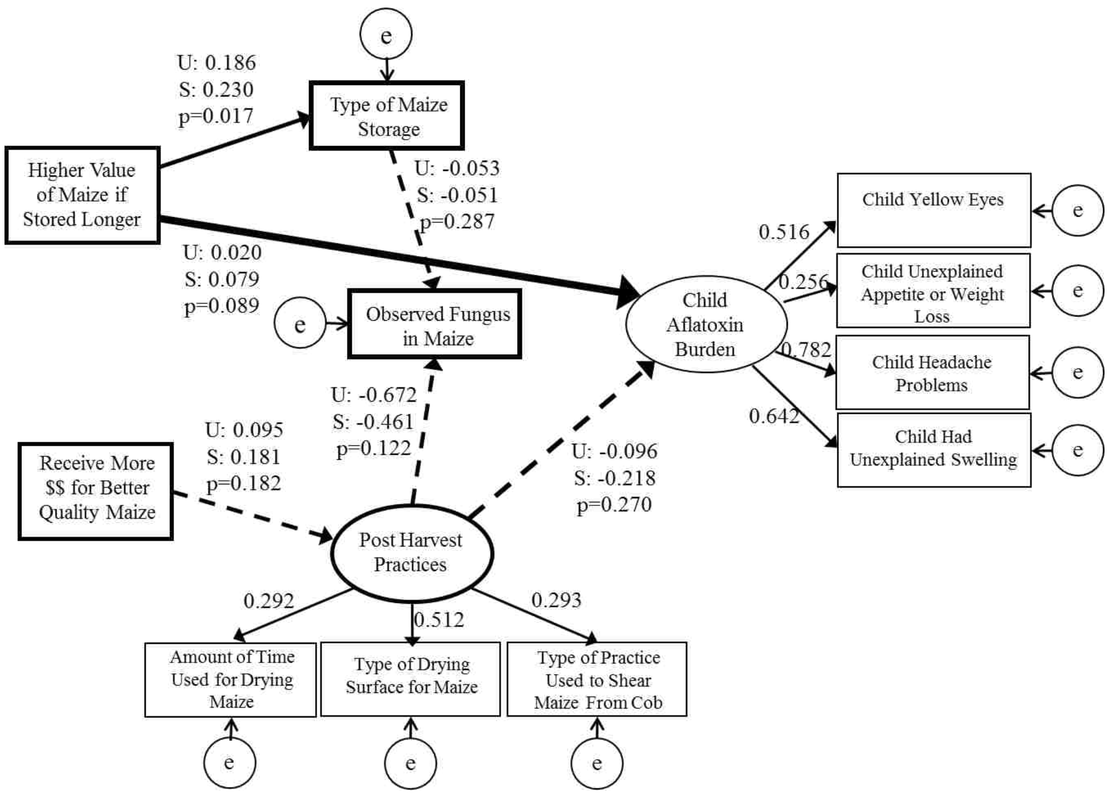


Figure 5. Final structural equation model of February 2017 subsistence model. DWLS robust estimator used; n = 160; Chi-square: 51.497, p=0.057; Robust RMSEA: 0.039 (CI: 0.000 – 0.063); Robust CFI: 0.773; Robust TLI: 0.668.

4. DISCUSSION

Table 4 summarizes the effect sizes of variables on Child Aflatoxin Burden for all models. Additionally, the total effect sizes that account for indirect effects created by mediating variables are reported.

Table 4. The direct effects are summarized from the depicted SEMs and the total effect sizes presented for both models, subsistence and maker, for both October 2016 and February 2017.

Variables	Direct Effect 2016	Total Effect 2016	Direct Effect 2017	Total Effect 2017
Market Model				
Purchase Habits	-0.220	-0.220	-	-
Market Remoteness	-	-	-	-
Observed Fungus	-0.125	-0.125	-0.174	-0.174
Subsistence Model				
Maize Storage	-0.082	-0.128	-	-
Observed Fungus	-0.111	-0.111	-	-
Post-Harvest Practices	-0.158	-0.142	-	-
Storage Profit	-0.068	-0.076	-0.051	-0.051
Quality Profit	-	-0.066	-	-

The results from the relative risk and odds ratio showed that a household who acquires their maize from the market has three times the odds (or is more than twice as likely) to have high levels of aflatoxins in their household maize as compared to subsistence households. This supports the hypothesis that the climatic environment is important for fungal growth as the ideal temperature for growth of Aspergillus is 23.0C to 26.0C (LIzarraga-Paulin et al., 2010), while reported temperatures in San Vicente are near 5.1C to 17.0C. Additionally, during informal interviews with local leaders it was reported that the two primary origins of the market maize were the southwest coast of Guatemala and Southern Mexico. Previous studies have reported high levels of aflatoxins in maize grown in these two regions where high temperatures and high humidity promote the growth of Aspergillus (Torres et al., 2015).

4.1. Market Maize

The final SEM for households who reported attending the market in October 2016 showed that purchase habits reported by the mother were significantly correlated with the Child Aflatoxin Burden latent variable. The habits were ranked based on the hypothesized decrease in probability of aspergillus growth and aflatoxin exposure; for example, the best answer possible was looking for fungus in the maize. Observed Fungus was negatively correlated with Child Aflatoxin Burden which was counter to the original hypothesis. During the collection of the maize samples, trained field workers looked for fungus within the maize storage area. This meant that the households where fungus was observed in the storage had fewer reported symptoms related to aflatoxin exposure among children. The counter result may have been due to the ability of the household to avoid maize when fungus was observed, therefore reducing putative aflatoxin exposure to their children.

Neither Purchase Habits nor Remoteness of Market were correlated with Observed Fungus. This meant that the association between Purchase Habits and Child Aflatoxin Burden was not mediated by trained staff observing fungus in the household maize. Furthermore, Remoteness of Market did not have a significant correlation with either Child Aflatoxin Burden nor Purchase Habits. ‘Remoteness’ was assessed by the time it took to drive from the primary regional distribution hub for maize (Quetzaltenango, Guatemala) to each market. The hypothesis was that the more ‘remote’ the market the higher the chance of aflatoxin presence. Bruns 2003 showed that the longer the transport time the higher the level of aflatoxin in the maize (Bruns, 2003). However, the correlations among the data suggested that this was insignificant in this location or at this time of year.

The final SEM for households who reported attending the market in February 2017 showed that only Observed Fungus had a significant correlation with Child Aflatoxin Burden. The month of February is two months after the harvest season in both the highlands and lowlands. In comparison, October is eleven months after the harvest season in the highlands and two months after the harvest season in the lowlands. This may have influenced the importance of purchase habits at the market. Therefore, if there are elevated levels of aflatoxins at the market, Purchase Habits may become more significant.

4.2. Subsistence Maize

Although the climate within San Vicente was not ideal for Aspergillus growth, both informal reports of problems with fungus in local maize crops and tested aflatoxic levels of 0-96 ppb suggested aflatoxins could be a potential issue within the region.

Post-Harvest Practices, Observed Fungus, Maize Storage, and Storage Profit were all confirmed as statistically correlated with Child Aflatoxin Burden in October 2016. As Post-Harvest Practices, Maize Storage, and Storage Profit increased or improved the Child Aflatoxin Burden reported by the mother decreased, supporting these original hypotheses. Post-harvest practice methods have attracted significant research as well as financial investments for interventions and farmer trainings (Wu & Khlangwiset, 2011). This study's findings further support these research aims and practitioner investments. Similarly, recent studies report correlations between improved storage of maize and the reduced level of aflatoxin within the maize (Chulze, 2010; Hell et al., 2008). Finally, Storage Profit was also correlated with Child Aflatoxin Burden. This correlation has had less focus among researchers interested in aflatoxin interventions, but may warrant further research if market-based interventions are of interest to implementing agencies. Additionally, a cost-benefit analysis would help identify the monetary return on investment and the population coverage per dollar spent for all potential interventions.

Observed Fungus had a negative correlation with Child Aflatoxin Burden which was counter to the original hypothesis, but supported the finding from the market maize SEM. Potentially, if the fungus was visible, it was possible to avoid consumption. Aflatoxin in maize can be difficult to detect as it can grow inside damaged kernels and therefore go undetected unless specific equipment is utilized.

Maize Storage and Post-Harvest Practices had statistically significant correlations with Observed Fungus. Maize Storage was negatively correlated with Observed Fungus suggesting that among households with improved storage practices, fungus was observed more often in the maize. This was counter to the original hypothesis. Potential explanations include; 1) an intricate relationship between the material used for maize storage and different types of species of fungal growth or 2) spurious correlation. Improved post-harvest practices for maize was correlated with a lower prevalence in observations of fungus among households, supporting the original hypothesis. Hell et al have reported correlations between several types of improved post-harvest practices and a reduction in aflatoxin presence (Hell et al., 2008). The three post-harvest practices used as indicators for this latent variable included the drying time of maize, the drying surface used, and the mechanisms used to remove maize kernels from the cob. These have been a focus for USAID in Central America and Sub-Saharan Africa.

Lastly, two questions were asked which were related to the relationship between physical barriers to fungal growth and economic incentives for farmers. Results suggested that if the household perceived receiving more money from buyers at the market as they increased the time they waited to sell their maize after harvest (higher maize price for longer storage) they would also have improved maize storage facilities. This supported the original hypothesis. The second question inquired if the household received more money for maize if it was of higher quality (Quality Profit) which was hypothesized to affect the type of post-harvest practices. A significant positive correlation was found meaning that those households who received more money for better quality maize utilized better post-harvest practices (specifically, drying time, drying surface, or shucking). Table 5 shows the total effect of both of these variables on Child Aflatoxin Burden and that Storage Profit had the largest total effect size. Interventions aimed both at distributing market price information as well as improving buyer recognition in fungal devaluation may provide options for reducing aflatoxin exposure to children.

For the subsistence SEM from February 2017, two months after harvest, only two correlations were significant; Storage Profit with Child Aflatoxin Burden and Storage Profit with Maize Storage. Similar to the market-based SEMs, the data suggest a large decrease in significant relationships between October 2016 and February 2017. However, significant correlations which remained over seasonal changes include Storage Profit on Child Aflatoxin Burden and Storage Profit on Maize Storage. This suggests that further research on potential effectiveness among interventions associated with maize storage or market price information may have value in reducing putative child aflatoxin exposure.

In this study data from the town of San Vicente was analyzed to assess hypothesized correlations between aflatoxin transmission and exposure from maize from local markets and subsistence farming. An odds ratio and relative risk ratio confirmed the hypothesis of a higher risk of putative AFB exposure among households attending local markets for maize acquisition. The SEM for market purchased maize in October 2016 confirmed that the purchase habits related to fungus awareness in maize was significant. The SEM for subsistence maize from October 2016 confirmed that the post-harvest practices, type of maize storage, and profit for storage were significant. Additionally, in both October 2016 and February 2017 counter to the original hypothesis, observed fungus in the maize storage was correlated with a decrease in reported child aflatoxin exposure symptoms. Finally, a higher number of significant variables correlated with Child Aflatoxin Burden were found one month before harvest as compared to two months after harvest. Because of the wide spread problem that aflatoxin presents it is critical for practitioners and policy makers to understand the complex relationships and potential intervention points.

REFERENCES

Agriculture and Consumer Protection. (n.d.). Household Food Security and Crop Diversification Among Smallholder Farmers in Guatemala. Rome, Italy. Retrieved from http://www.fao.org/docrep/U8050T/U8050T06.HTM.

Bbosa, G. S., Kitya, D., Lubega, A., Ogwal-Okeng, J., Anokbonggo, W. W., & Kyegombe, D. B. (2013). Chapter 12: Review of the Biological and Health Effects of Aflatoxins on Body Organs and Body Systems. In Aflatoxins - Recent Advances and Future Prospects (pp. 239–265).

Bruns, H. A. (2003). Controlling Aflatoxin and Fumonisin in Maize by Crop Management. Toxin Reviews, 22(2–3), 153–173. https://doi.org/10.1081/TXR-120024090.

Chulze, S. N. (2010). Strategies to reduce mycotoxin levels in maize during storage: a review. Food Additives & Contaminants. Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment, 27(5), 651–657.

Daniel, W. W. (1995). Biostatistics: A Foundation for Analysis in the Health Sciences.

European Commission. Commission Regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foodstuffs (2006). https://doi.org/10.2203/dose-response.06-012.Hanekamp.

Grace, J. B. (2006). Structural Equation Modeling and Natural Systems (First). United Kingdom: Cambridge University Press.

Hell, K., Fandohan, P., Bandyopadhyay, R., Kiewnick, S., Sikora, R., & Cotty, P. J. (2008). Pre- and postharvest management of aflatoxin in maize: An African perspective. Mycotoxins: Detection Methods, Management, Public Health and Agricultural Trade, 209–218. https://doi.org/10.1079/9781845930820.0219.

International Agency for Research on Cancer. (2006). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Aflatoxins.

LIzarraga-Paulin, E. G., Moreno-Martinez, E., & Miranda-Castro, S. P. (2010). Chapter 13: Aflatoxins and their Impact on Human and Animal Health: An Emerging Problem. In Aflatoxins - Biochemistry and Molecular Biology (pp. 1–28).

Mîndrilã, D. (2010). Maximum likelihood (ML) and diagonally wighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. International Journal of Digital Society, 1(1), 60–66. https://doi.org/10.20533/ijds.2040.2570.2010.0010.

Neogen Corporation. (2012). The Quantification of Aflatoxins in Foods. Lansing, MI. Retrieved from http://www.neogen.com/foodsafety/pdf/procedures/.

Torres, O., Matute, J., Gelineau-van Waes, J., Maddox, J. R., Gregory, S. G., Ashley-Koch, a. E., … Riley, R. T. (2015). Human health implications from co-exposure to aflatoxins and fumonisins in maize-based foods in Latin America: Guatemala as a case study. World Mycotoxin Journal, 8(2), 143–159. https://doi.org/10.3920/WMJ2014.1736.

U.S. Food and Drug Administration. Food and Drug Administration Compliance Program Guidance Manual: Mycotoxins in Domestic and Imported Foods (2016). https://doi.org/10.1300/J138v04n01_14.

United Nations Children's Emergency Fund. (2013). Improving Child Nutrition: The achievable imperative for global progress. Retrieved from http://www.unicef.org/gambia/Improving_Child_Nutrition_-_the_achievable_imperative_for_global_progress.pdf.

United Nations General Assembly. Transforming our world: the 2030 Agenda for Sustainable Development, A/70/L.1 § (2015).

United States Agency International Development. (2012). Guatemala Country Development Cooperation Strategy 2012-2016.

United States Agency International Development. (2014). Baseline Study for the Title II Development Food Assistance Programs in Guatemala.

Voth-Gaeddert, L. E., Stoker, M., Torres, O., & Oerther, D. B. (n.d.). Analysis of Correlations between Aflatoxin B, Intestinal Function, and Child Height-for-Age Among Young Children in Guatemala. International Journal of Epidemiology.

Wild, C. P., Miller, J. D., & Groopman, J. D. (2016). Mycotoxin Control in Low- and Middle- Income Countries.

Wu, F., & Khlangwiset, P. (2011). Health economic impacts and cost-effectiveness of aflatoxin reduction strategies in Africa: Case studies in biocontrol and postharvest interventions, 1–21. https://doi.org/10.1080/19440040903437865.Health.

V. ACUTE TO CHRONIC MALNUTRITION: HOW SIGNIFICANT WATER, SANITATION, AND HYGIENE FACTORS CHANGE WITH HEALTH OUTCOMES AND GEOGRAPHIES IN THE WESTERN HIGHLANDS OF GUATEMALA

To be submitted to the journal of Environmental Science and Technology

Voth-Gaeddert LE1*, Oerther DB1

1 Civil, Architectural, and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO, USA, lv6w3@mst.edu

*Corresponding Author

ABSTRACT

Adequate and appropriate water, sanitation, and hygiene (WaSH) infrastructure is important for reducing pathogen exposures in developing communities. To improve the ability of field practitioners to optimize WaSH infrastructure systems within communities, developing models can provide insight to the complex interactions among WaSH infrastructure, health outcomes, and geographies. This study investigates the significant correlations among WaSH infrastructure variables and three different health outcomes (diarrhea, environmental enteric dysfunction, and stunting) over five geographic regions within Guatemala. Exploratory structural equation modeling was used to build WaSH models from US Agency for International Development (USAID) 2012 Food for Peace Survey data (n=2,103). Validity of the models was then tested utilizing the USAID 2013 Western Highlands Integrated Program survey data collected from the same regions (n=4,633). Results confirmed the original hypothesis that significant WaSH infrastructure variables widely vary over health outcome and geographic region. A nested relationship was found between the 2012 models and 2013 models partially supporting the validity of the models. The 'floor' pathogen transmission pathway was identified as significant across all geographies for child stunting. Additionally, commonalities in potential pathogen transmission pathways were identified among environmentally similar geographies. Practitioners and policy makers must account for specific health outcomes and understand which WaSH infrastructure intervention is most appropriate at the correct scale.

Keywords: Water, Sanitation, Hygiene, Infrastructure, Diarrhea, Child Stunting, Environmental Enteric Dysfunction

1. INTRODUCTION

Adequate coverage of water, sanitation, and hygiene (WaSH) infrastructure in Central America has been reported to be low as compared to overall Latin American averages (Uytewaal 2016). However, previous research has reported on the positive significance of WaSH interventions for the health of communities in these regions (Fewtrell et al. 2005; Moll et al. 2007). A primary objective for WaSH infrastructure in developing countries is to create barriers to transmission of bacterial contaminants from one person or animal to another person. These transmission pathways have previously been summarized as the ‘five Fs’; fingers, fluids, floors, foods, and flies (Center for Disease Control and Prevention 2013; The World Bank 2014). Due to the variety of pathogen species, the differing severities of exposure, repeated exposures, and the impact on intestinal integrity of children; the understanding of the relationships between WaSH infrastructure barriers and health outcomes is limited (Waddington et al. 2009). Additionally, effectiveness of WaSH infrastructure on improving health outcomes has been shown to be geographically dependent, as moving from one community or region to another may alter coverage rates, environmental realities, or cultural interactions (Botting et al. 2010).

The US Agency for International Development (USAID) consistently collects household WaSH infrastructure data which includes water sources, water treatment techniques, types of sanitation facilities, presence of soap at hand washing stations, and floor type or animal pen infrastructure. Furthermore, USAID collects specific child health data including child stunting, child wasting, child body-mass-index, and diarrheal occurrences (United States Agency International Development n.d.). Child stunting (or wasting) is defined as a child with a height-for-age (or weight-for-height) ratio two standard deviations below the World Health Organization growth mean and is often used as a chronic (or acute) health indicator (World Health Organization 2010). Presence of diarrhea is often used as an acute measure of health and is defined by the WHO as three or more loose stools in 24 hours (World Health Organization 2017). With regional WaSH infrastructure data coupled with health data, it is possible to assess trends over geographies and health outcomes to help identify significant infrastructure-based interventions that may have the best return on investment for improving child health.

To study the geographic and health latency based correlations of various WaSH infrastructure and health outcomes in the western highlands of Guatemala, two datasets from USAID Guatemala were assessed. Structural equation models were built and tested over five geographic regions and three latencies of health. Results supported the original hypothesis in that the types of WaSH infrastructure that correlated with specific health outcomes were dependent on both the geographic location and health outcome latency (acute to chronic).

2. METHODS

2.1. Location

Data from both the USAID 2012 Food for Peace Baseline Survey (United States Agency International Development 2014) and USAID 2013 Western Highlands Integrative Program Baseline Survey (United States Agency International Development & Measure Evaluation 2014) were collected via household visits within 30 municipalities (counties) in five departments (states) of Guatemala. Household visits included an orally administered survey given to the mother in the local dialect and anthropometric measurements of the child (following WHO protocol). Both surveys were randomized cluster samples with the 2013 survey sampling population being expanded to include more children within the same municipalities (United States Agency International Development 2014; United States Agency International Development & Measure Evaluation 2014). Data was collected with the approval of the Ministry of Health and consent for analysis of the deidentified data was attained from USAID.

The departments of Guatemala included Huehuetenango, San Marcos, Quiche, Totonicapán, and Quetzaltenango. Table 1 reports environmental statistics on each department including mean elevation, mean temperatures, and mean rainfall. All five departments are in a set of mountain ranges collectively known as the western highlands. Commonalities among the population included 1) farming as the primary livelihood and 2) the level of socio-economic status with over 51% of the population lives below the poverty line (United States Agency International Development 2012; Prado Córdova et al. 2013). A majority of the population self-identified as a specific Mayan ethnicity including Ixil, Quiché, Mam, and Popti each utilizing their own distinct language (United States Agency International Development 2014).

Table 1. Environmental statistics for each department; elevation in meters, mean temperature span over the year in celsius, and annual rainfall in millimeters.

Department	Elevation (m)	Mean Temperature (C)	Annual Rainfall (mm)
Huehuetenango	2000-2700	17.9-20.9	2700
San Marcos	2400-2600	10.0-16.0	1450
Quiche	1600-1850	13.0-18.0	1500
Totonicapán	2100-2500	6.5-13.1	800-1200
Quetzaltenango	2100-2500	9.5-14.2	1300

2.2. Data Preparation

Table 2 shows the variables selected to be analyzed in the models along with the associated questions and scales used. Diarrhea and ZHAZ (height-for-age z-score; child stunting metric) were selected as acute and chronic measures of health, respectively, while the latent variable EED was created to represent medium-term measures of health. Latent variables are discussed below, but indicator variables used for EED included ZHAZ, ZBMI (body-mass index z-score), ZWHZ (weight-for-height z-score), and Diarrhea. All WaSH scales are perceived to get worse as they increase, while all health outcomes are perceived to get better as they increase. Additionally, based on prior research findings each WaSH infrastructure variable was associated with the five-f transmission pathway(s) in which it provided a barrier for (Julian 2016; Prüss et al. 2002; Center for Disease Control and Prevention 2013; The World Bank 2014). WaterSource and WaterTreat infrastructure were associated with barriers of transmission via the fluids and foods pathways. HygSoap was associated with barriers for the fingers and foods transmission pathway. SanitType was associated with barriers for transmission for floors and flies' pathways. Finally, AnimalPen/FloorType were associated with barriers for the floor transmission pathway.

Table 2. Variables, explanations, and scales used in the structural equation models

Variable	Explanation	Scale
Diarrhea	Has the child had a bout of diarrhea within the past two weeks?	0 = yes; 1 = no
EED	Latent [reflective] variable, created from the manifestations of Diarrhea, Height-for-Age z-score (ZHAZ), Body Mass Index z-score (ZBMI), and Weight-for-Height z-score (ZWHZ)	ZHAZ, ZBMI, and ZWHZ given in standard deviations
ZHAZ	Measure of height to age of child and standardized based on World Health Organization growth charts	Given in standard deviations
WaterSource	What water source is used by the household?	1 = In house connected to system 2 = Out of house connected to system 3 = Public tap 4 = Private pump 5 = Public pump 6 = River, lake, open water 7 = rainwater 8 = Pick-up truck tank
WaterTreat	Does the household treat their water in anyway?	0 = yes; 1 = no
HygSoap	Was soap observed at the households' handwashing station?	0 = yes; 1 = no
SanitType	What type of sanitation facility is used by the household?	1 = In house connected to system 2 = In house connected to septic tank 3 = Latrine 4 = Open latrine/hole 5 = No sanitation facility
AnimalPen	Does the family have an animal pen that has walls?	0 = yes; 1 = no

The only discrepancy among variables used in the models was the variable selected to represent the 'floor/field' enteric disease transmission pathway. The 2012 dataset included the question associated with 'AnimalPen' which was selected to represent the floor/field transmission pathway in the model based on previous evidence which suggested an increase in free roaming animals near the house increased the probability of enteric disease exposure via the floor/field for children (Zambrano et al. 2014). The 2013 dataset included the question associated with 'FloorType' which was selected as the substitute for AnimalPen (not collected in 2013) based on the hypothesis that the quality of floor was associated with the probability of exposure to the child via the floor transmission pathway (Douglas S et al. 2002).

2.3. Statistical Techniques

Three structural equation models (SEM) were built and tested for five geographic regions and each model included five WaSH infrastructure variables (WaterSource, WaterTreat, SanitType, HygSoap, and AnimalPen or FloorType) regressed on by a health variable (Diarrhea, EED, or HAZ). SEM is a statistical modeling technique which combines path analysis and factor analysis to analyze multiple hypotheses simultaneously. Figure 1 depicts the basic graphical representation of an SEM where arrows are hypotheses, rectangles are observable variables, and ovals are latent variables. A latent variable (shown here as 'EED') is hypothesized to be an underlying factor which influences a set of indicator variables (shown here as 'ZHAZ', 'ZBMI', 'ZWHZ', and 'Diarrhea'). As this factor is estimated, path analysis is used to compute and analyze the difference in the data driven and hypothesized covariance matrices. These covariance matrices include all observable and latent variables. If the data show good fit to the model based on four fit statistics (Chi-square $p>0.05$, Root Mean Square Error of Approximation<0.08, Confirmatory Factor Index>0.90, Tucker Lewis Index>0.90) the individual parameter estimates can be analyzed (read like regression parameter estimates). An exploratory SEM approach was used to build the models from the 2012 data while a confirmatory approach was utilized to test the validity of each model using the 2013 data. The Lavaan package in R 3.3.2 was used for the analysis. Further reading on SEM is encouraged and can be found in Grace 2006. (Grace 2006).

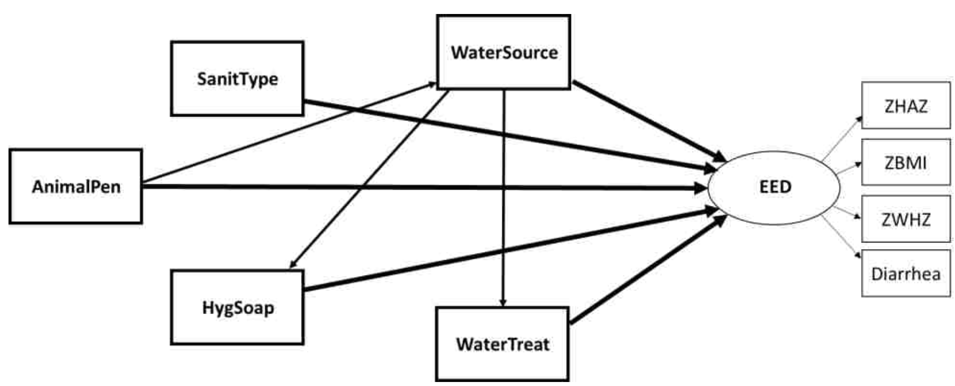


Figure 1. A hypothesized WaSH infrastructure structural equation model with the EED outcome variable. Arrows are hypothesized causality, rectangles are observable variables, and ovals are latent variables.

Finally, utilizing previously reported transmission pathways associated with individual WaSH infrastructure barriers (discussed above; 5Fs), potential transmission pathways were identified for each geography and health outcome based on the set of

2012 SEMs. Additionally, both changes in diarrheal occurrences and stunting levels between the 2012 and 2013 datasets are reported alongside the changes in WaSH infrastructure based transmission pathways.

3. RESULTS

3.1. Descriptive Results of Data

Table 3 presents descriptive statistics for each dataset. The 2012 data included n=2,103 children included in the analysis, 52% males and 48% females. Diarrheal prevalence within the past two weeks was 39% and the child stunting level was -2.47 SDs. The 2013 data included n=4,633 children included in the analysis, 51% males and 49% females. Diarrheal prevalence within the past two weeks was 33% and the child stunting level was -2.44 SDs. Data was grouped according to geographic proximity of each municipality which resulted in three separate departments, Huehuetenango, San Marcos, and Quiche, and two sub-divided departments, Northern Totonicapán and Quetzaltenango-Southern Totonicapán. According to the data, the diarrheal prevalence improved in every group from 2012 to 2013, while child stunting became worse in every group except San Marcos over the same time period. The 2013 sample size for Huehuetenango, San Marcos, and Quiche was over double the 2012 dataset, while Northern Totonicapán and Quetzaltenango-Southern Totonicapán retained similar sample sizes.

Table 3. Descriptive statistics for both USAID 2012 and USAID 2013 datasets

	USAID 2012 Dataset	USAID 2013 Dataset
Sample Size	2,103	4,633
Boys	52%	51%
Girls	48%	49%
Diarrhea Prevalence	39%	33%
Child Stunting Level	-2.47 standard deviations	-2.44 standard deviations

3.2. 2012 Model Results

Figure 2 displays the graphical results of the set of SEMs built by the 2012 data and tested using the 2013 data for San Marcos. Table 4 presents results for all groups on the significant WaSH infrastructure variables (at a 10% level) identified by the 2012 models for acute (diarrhea), medium (EED), and chronic (ZHAZ) health outcomes. Standardized parameter estimates are also reported to provide a rank order for variables.

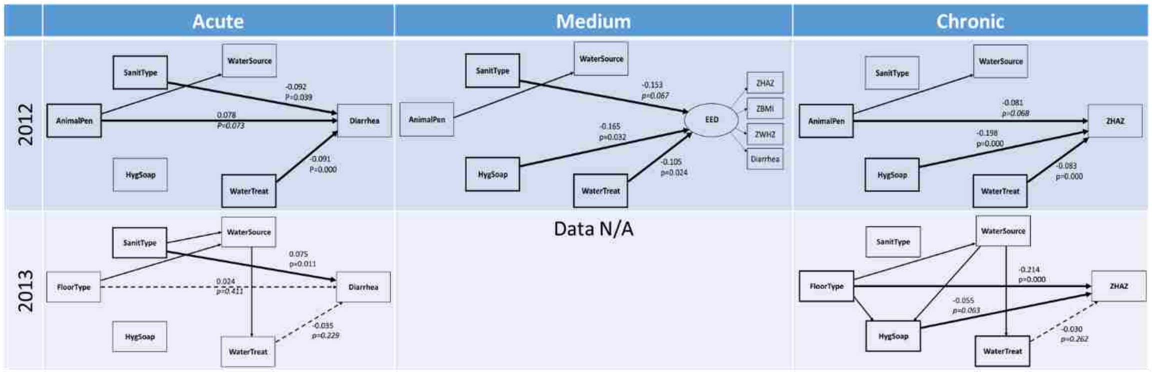


Figure 2. The set of five structural equation models for the department of San Marcos. Outcome variables include Acute (Diarrhea), Medium (EED latent variable), and Chronic (child height-for-age). 2012 denotes models built from the USAID 2012 dataset and 2013 are the results of the test of the 2012 models with the USAID 2013 dataset.

From Table 4 the Huehuetenango models had no significant WaSH infrastructure variables correlated with Diarrhea (acute), however, in both EED (medium) and ZHAZ (chronic) models SanitType was negatively correlated to the health outcome. Furthermore, for the ZHAZ model, WaterSource and HygSoap were also negatively correlated with the health outcome. For the San Marcos models, WaterTreat was negatively correlated to the health outcome in all models, SanitType was negatively correlated with Diarrhea and EED, and AnimalPen was positively correlated with Diarrhea and negatively correlated with ZHAZ. Additionally, HygSoap was negatively correlate with EED and ZHAZ. The Quiche models only had WaterSource (positively correlated) and WaterTreat negatively correlated with Diarrhea. For the Northern Totonicapán models, SanitType was negatively correlated to the health outcome in all models and AnimalPen was negatively correlated with ZHAZ. Finally, the Quetzaltenango-Southern Totonicapán models had WaterSource negatively correlated with Diarrhea, AnimalPen positively correlated with EED, and SanitType negatively correlated with ZHAZ.

Table 4. Summarized results of the 2012 Models for all health outcomes and all geographic groups. The proxy for the acute outcome was diarrhea, medium was EED, and chronic was child height-for-age.

Region	Acute	Medium	Chronic
Huehuetenango	-	SanitType (-0.220)	WaterSource (-0.152) SanitType (-0.148) HygSoap (-0.118)
San Marcos	SanitType (-0.092) WaterTreat (-0.091) AnimalPen (0.078)	HygSoap (-0.165) SanitType (-0.153) WaterTreat (-0.105)	HygSoap (-0.198) WaterTreat (-0.083) AnimalPen (-0.081)
Quiche	WaterSource (0.211) WaterTreat (-0.111)	-	-
Northern Totonicapán	SanitType (-0.108)	SanityType (-0.228)	AnimalPen (-0.140) SanitType (-0.101)
Quetzaltenango + Southern Totonicapán	WaterSource (-0.091)	AnimalPen (0.145)	SanitType (-0.085)

3.3. 2013 Model Results

In Figure 2, the 2013 row displays the SEMs graphically for San Marcos. Data was not available for computing the EED models. Furthermore, AnimalPen was not available in the 2013 dataset and was therefore replaced with FloorType. The results of the confirmation analysis failed to show exact fit of the 2013 data to the 2012 models. Table 5 reports the changes (additions and subtractions) to each model to attain adequate fit of the model to the data. For each model, minimal adjustments were made to attain fit of the 2013 data to the 2012 model according to the tests of model fit as discussed above. While full validation via model fit statistics was not attained, the adjusted 2013 models (see Figure 2 for San Marcos example) demonstrated a parenting effect, with the 2012 models being nested (a sub-model) within the 2013 models.

Over the 2012-2013 period the diarrheal prevalence among the population of children surveyed in Huehuetenango dropped 8.1%, while child stunting became worse by 0.01 SD. From Table 5 in Huehuetenango WaterSource became negatively correlated with Diarrhea while FloorType (replacement for AnimalPen) became negatively correlated with ZHAZ. The diarrheal prevalence for the study population in San Marcos dropped 12.6% and child stunting improved 0.22 SD. FloorType and WaterTreat became insignificant with Diarrhea while WaterTreat became insignificance with ZHAZ. The diarrheal prevalence for the study population in Quiche dropped 1.8% and child stunting worsened by 0.1 SD. WaterTreat became negatively correlated with Diarrhea while WaterSource became positively correlated and FloorType became negatively correlated with ZHAZ. The diarrheal prevalence for the study population in Northern Totonicapán dropped 14.5% while child stunting became worse by 0.18 SD. Within Northern Totonicapan several variables became important to model fit, but only HygSoap became negatively correlated with ZHAZ. Finally, the diarrheal prevalence for the study population in Quetzaltenango-Southern Totonicapán dropped 5.8% while child stunting became worse by 0.06 SD. FloorType and HygSoap also became negatively correlated with ZHAZ.

Table 5. Summarized results of the adjustments necessary for fit of the USAID 2013 dataset to the 2012 Models. *Italicized* names are variables that became insignificant in 2013. *Asterisked* names are variables that became important to the model but were not individually significant.

Regions	Acute	Chronic
Huehuetenango	WaterSource (-0.056; p=0.032) *SanitType* (0.030; p=0.275)	FloorType (-0.207; p=0.000)
San Marcos	*FloorType* *WaterTreat*	*WaterTreat*
Quiche	*WaterTreat*	WaterSource (0.177; p=0.000) FloorType (-0.156; p=0.000)
Northern Totonicapán	*WaterSource* (0.090; p=0.115) *FloorType* (-0.077; p=0.192)	HygSoap (-0.124; p=0.036) *WaterTreat* (0.091; p=0.162)
Quetzaltenango + Southern Totonicapán	-	FloorType (-0.188; p=0.001) HygSoap (-0.137; p=0.001)

3.4. Transmission Pathways

Finally, Figure 3 displays which potential pathogen transmission pathways were important for each geographical region and health outcome based on the significant WaSH infrastructure variables identified in the 2012 SEMs. In Huehuetenango no variables had a significant correlation for the diarrhea health model. However, type of sanitation facility was correlated with both EED and ZHAZ suggesting a barrier for the transmission pathway of floors and/or flies was important for both the medium and chronic health of the child. Furthermore, the type of water source and soap present at the hand washing station were significant for ZHAZ suggesting the transmission pathway of foods, fluids and/or fingers were additionally important to the chronic health of the child.

In San Marcos, significant correlations between the WaSH infrastructure variables and health outcomes suggested the following transmission pathways may have been important. For the acute health issue; fluids, foods, floors, and flies; for medium health issues; all pathways may have been important; and for the chronic health issue; fluids, fingers, foods, and floors. In Quiche, only the diarrhea health model had significant

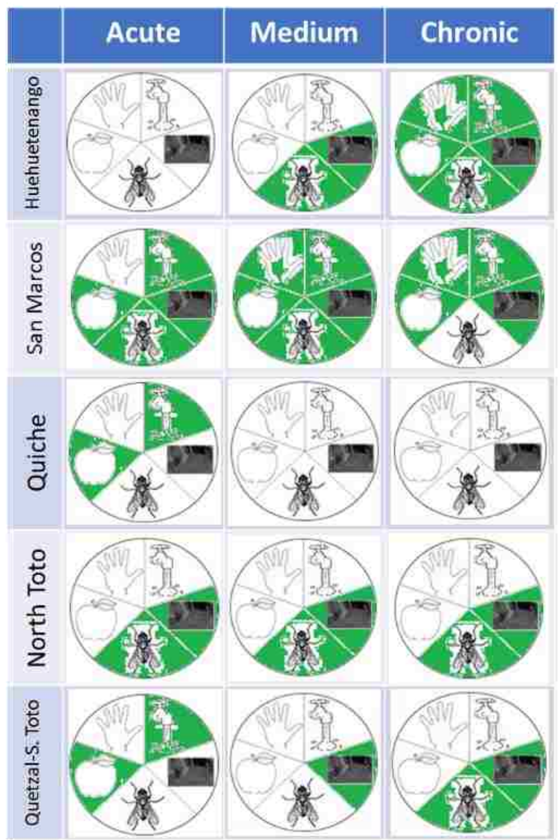


Figure 3. A summary of the potential transmission pathways (the 5Fs) that are causing problems based on the regional SEMs from 2012. The hand is fingers, water faucet is fluids, the ground is the floor, the fly is flies, and the apple is foods.

correlations with WaSH infrastructure variables which included the type of water source and type of water treatment technique suggesting the fluids and/or foods transmission pathway may have contributed to acute health problems. In Northern Totonicapán, type of sanitation facility was significant for all health outcomes which has been shown to be associated with the floor and/or fly transmission pathway. Additionally, owning an animal pen was significant for the ZHAZ health outcome suggesting the floor transmission pathway was important to chronic health. Finally, in Quetzaltenango-

Southern Totonicapán a different WaSH infrastructure variable was significant for each health outcome. For diarrhea, the type of water source suggested fluids and/or foods were important to acute health; for EED, animal pens suggested floors were important for medium health; and for ZHAZ, type of sanitation suggested floors and/or flies were important for chronic health.

Figure 4 depicts the changes within each geography and health outcome in the 2013 dataset (orange; became insignificant, light green; became significant). Furthermore, it gives the direction and magnitude of change in the specific health indicator (diarrhea, DIA; height-for-age, HAZ). In Huehuetenango, the type of water source and type of floor became correlated with the diarrhea and ZHAZ health outcomes of the child, respectively. This meant that for the transmission pathways that affected acute health, fluids and foods were potentially significant while floors were already important in the chronic transmission pathways. In San Marcos, several variables became insignificant in regards to the parameter estimates but remained important to the overall models. Type of floor and type of water treatment technique, associated with floors, fluids, and foods, were not correlated with diarrhea in the 2013 model. Type of water treatment technique also became insignificant with ZHAZ in the chronic health model. In Quiche, again, type of water treatment technique became insignificant in the diarrhea health model, but all significant transmission pathways from 2012 remained important due to other significant WaSH variables. In the ZHAZ model the type of water source and type of floor became significant suggesting the fluids, foods, and floors transmission pathways became important. In Northern Totonicapán, only the presence of soap at the handwashing station became correlated in the ZHAZ model which suggested that the fingers and foods transmission pathways were important for chronic health. Finally, in Quetzaltenango-Southern Totonicapán type of floor and presence of soap became correlated in the ZHAZ model. Accounting for previous correlations of other WaSH infrastructure variables, only fingers and foods appear to have become significant for chronic health.

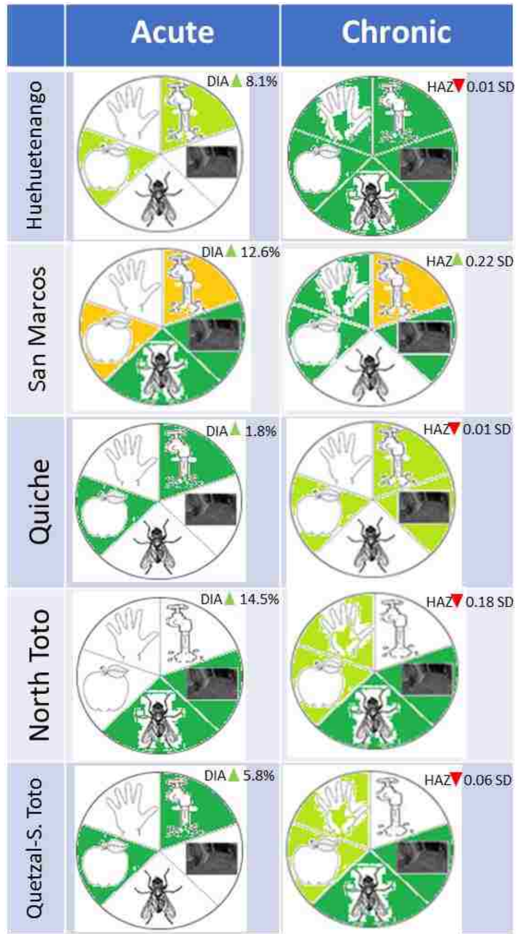


Figure 4. A summary of the potential transmission pathways (the 5Fs) that are causing problems based on the regional SEMs from 2013. Green pathways are confirmed pathways from 2012, light green pathways are new potential pathways, and orange pathways are green pathways from 2012 which became insignificant. The hand is fingers, water faucet is fluids, the ground is the floor, the fly is flies, and the apple is foods.

4. DISCUSSION

4.1. 2012 Models

For the 2012 SEMs, each geographical group displayed a unique set of significant WaSH infrastructure variables which also changed for each health outcome. Overall, SanitType was the most common significant WaSH variable among the 15 models being significant in 8. Having soap at the handwashing station was correlated with medium and/or chronic health outcomes in Huehuetenango and San Marcos (3 of 4 models), while improved water sources were important for acute outcomes in Totonicapán and Quetzaltenango. A common variable across all health outcomes for San Marcos was the type of water treatment and in Northern Totonicapán type of sanitation was common across all health outcomes.

Figure 3 displays the results of the 5F transmission pathways associated with the identified WaSH variables from the 2012 SEMs. If a pathway was important (green in Figure 3) this could potentially mean that 1) there was a high number or a longer sustained level of pathogens transmitted via this particular pathway which made the associated WaSH barrier significant, 2) there was a wide enough distribution of exposure levels for a given pathway and barrier effectiveness to be correlated with a health outcome, or 3) there was a spurious correlation. In Figure 3, an analysis of individual columns provides a comparison across geographic regions. Models for Huehuetenango and San Marcos displayed similarities among potentially associated transmission pathways, while models for Northern Totonicapán and Quetzaltenango-Southern Totonicapán displayed similar characteristics. The similarities in transmission pathways regionally appeared to be more pronounced in the medium and chronic health indictors (EED and HAZ) as flies and floors were important for the Totonicapán-Quetzaltenango region while fingers, fluids, floors, and foods were all significant for the Huehuetenango-San Marcos region. In the acute health indicator column (diarrhea) the data suggested two potential geographic regions held similar transmission pathway characteristics. Models for San Marcos and Northern Totonicapán displayed a trend in flies and floors while models for San Marcos, Quiche, and Quetzaltenango-Southern Totonicapán had similarities in fluids and foods potentially contributing to diarrheal issues. The most common transmission pathway across all geographic groups and health outcomes was floors in the chronic health model column. This finding supports previous research on both increased levels of pathogens in this pathway as well as the quality of barriers for preventing transmission via the floor pathway (Zambrano et al. 2014; Douglas S et al. 2002; Al-Mazrou et al. 1995; Exum et al. 2016).

4.2. 2013 Models

The SEMs from 2013 demonstrated a nesting effect to the 2012 SEMs. An example of a nesting effect would be, Model A is said to be nested in Model B if Model A is a submodel of Model B, where Model B contains all significant correlations of

Model A, but Model A does not contain all significant correlations of Model B (Kline 2011). The testing of the 2012 models with the 2013 data did not provide exact model confirmation, but offered partial model validation due to the nested relationship the 2012 models shared with the 2013 models. This nesting effect may have been due to one of several factors. First, changes in the use of WaSH infrastructure within households over one year may change the dynamics of the relationship between WaSH and health. Second, the overall dataset doubled in sample size and for the regional groups of Huehuetenango, San Marcos, and Quiche, their 2013 sample sizes included two to four times the number of children. This inclusion of a broader set of children introduces the possibility of additional correlations being significant, while potentially retaining the original correlations, creating a nested effect. Finally, the substitution of FloorType for AnimalPen may have caused slight discrepancies between the models.

Comparing geographic groups, descriptive statistics show San Marcos had two differentiating features. First, both diarrheal prevalence and the mean child height-for-age score improved and, second, three total WaSH infrastructure variables become insignificant; two for diarrhea (FloorType and WaterTreat) and one for ZHAZ (WaterTreat). For the remainder of the geographic groups several trends were identified and can be seen in Table 5. First, the type of floor became significant for three of the five child stunting models, although this variable was a replacement for the presence of an animal pen variable. Additionally, in Totonicapán and Quetzaltenango having soap at the handwashing station became significant for child stunting. This could either mean that an increase in the presence of soap was correlated with a decrease in child stunting or that households improved the actual usage of the soap at the handwashing station.

An addition (or subtraction) of a significantly correlated WaSH variable in a given model could suggest 1) a change in the number of pathogens being transmitted via that pathway and therefore making that barrier more (or less) important, 2) the distribution of use in the types of barriers for one variable increased (or decreased) and therefore became more (or less) detectable for significance, or 3) a spurious correlation. The diarrheal prevalence in Quiche, Totonicapán, and Quetzaltenango improved while the potential transmission pathways remained the same. However, child stunting became slightly worse for the same groups while, according to the ZHAZ models, multiple transmission pathways may have become significant; most commonly fingers and foods. In Huehuetenango, the opposite trend was present, as the diarrheal prevalence dropped, the type of water source became significant and therefore fluids and foods were potentially contributing transmission pathways. Furthermore, ZHAZ stayed constant and the potential transmission pathways also remained the same, even though the WaSH variable, FloorType, became significant. Finally, within the ZHAZ models, all geographic groups showed either SanitType or FloorType as significant suggesting the floor pathway was common among all groups. Previous work in Totonicapán identified a negative correlation between the height-for-age of the child and number of times the child played, supporting the hypothesis that an important pathogen transmission pathway is the floor (Voth-Gaeddert et al. n.d.).

This study assessed two datasets covering five departments of the western highlands of Guatemala by building and testing descriptive models of WaSH infrastructure variables and different health outcomes. Results showed a nested relationship between 2012 models and 2013 models partially supporting the validity of the models. Furthermore, the floor pathogen transmission pathway was identified as potentially common across all geographic regions for child stunting and was supported by previous work in the western highlands. For policy makers and practitioners at the municipality or department level, attention should be given to the correlations between WaSH variables and varying health outcomes within specific geographic groups while policy makers and practitioners at the regional or national level should be concerned with similarities across geographies in the same health outcome. It is only by understanding trends across geographies and health outcomes of interest that change will be possible on a national scale within Guatemala.

REFERENCES

Al-Mazrou, Y.Y. et al., 1995. Factors Associated with Diarrhoea Prevalence in Saudi Arabia. Journal of Family Community Medicine, 2(1), pp.27–34.

Botting, M.J. et al., 2010. Water and sanitation infrastructure for health: The impact of foreign aid. Globalization and health, 6(1), p.12. Available at: http://globalizationandhealth.biomedcentral.com/articles/10.1186/1744-8603-6-12.

Center for Disease Control and Prevention, 2013. Diarrhea: Common Illness, Global Killer,.

Douglas S, B. et al., 2002. Effect of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late hildhood: a follow-up study. Lancet, 359, pp.564–71.

Exum, N.G. et al., 2016. Floors and Toilets: Association of Floors and Sanitation Practices with Fecal Contamination in Peruvian Amazon Peri-Urban Households. Environmental Science and Technology, 50(14), pp.7373–7381.

Fewtrell, L. et al., 2005. Water, sanitation, and hygiene interventions to reduce diarrhoea in less developed countries: a systematic review and meta-analysis. The Lancet infectious diseases, 5(1), pp.42–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15620560.

Grace, J.B., 2006. Structural Equation Modeling and Natural Systems First., United Kingdom: Cambridge University Press.

Julian, T.R., 2016. Environmental transmission of diarrheal pathogens in low and middle income countries. Environmental Science: Processes Impacts, 18, pp.944–955. Available at: http://xlink.rsc.org/?DOI=C6EM00222F.

Kline, R.B., 2011. Principles and Practice of Structural Equation Modeling 3rd ed., New York: The Guilford Press.

Moll, D.M. et al., 2007. Health impact of water and sanitation infrastructure reconstruction programmes in eight Central American communities affected by Hurricane Mitch. Journal of Water and Health, 5(1), pp.51–65.

Prado Córdova, J.P. et al., 2013. Rural income and forest reliance in highland guatemala. Environmental Management, 51(5), pp.1034–1043.

Prüss, A. et al., 2002. Estimating the Burden of Disease from Water, Sanitation, and Hygiene at a Global Level. Environmental Health Perspectives, 110(5), pp.537–542.

The World Bank, 2014. Health Benefits - Sanitation and Hygiene: Why They Matter. Available at: http://water.worldbank.org/shw-resource-guide/sanitation-and-hygiene-why-they-matter/health-benefits.

United States Agency International Development, 2014. Baseline Study for the Title II Development Food Assistance Programs in Guatemala,.

United States Agency International Development, 2012. Guatemala Country Development Cooperation Strategy 2012-2016,.

United States Agency International Development, The DHS Program: Demographic and Health Surveys. Available at: http://dhsprogram.com/.

United States Agency International Development & Measure Evaluation, 2014. The Western Highlands Integrated Program (WHIP) Evaluation Baseline Study in Guatemala: A Case Study in Evaluation Practice,.

Uytewaal, E., 2016. Situational Assessment of the Water, Sanitation and Hygiene sector in Latin America,.

Voth-Gaeddert, L.E. et al., What Causes Childhood Stunting Among Children of San Vicente, Guatemala: Employing Complimentary, System-Analysis Approaches. Journal of Environmental Health.

Waddington, H. et al., 2009. Water, Sanitation and Hygiene Interventions to Combat Childhood Diarrhoea in Developing Countries. 3ie Synthetic Reviews, (August), p.119.

World Health Organization, 2010. Country Profile Indicators: Interpretation Guide, Geneva, Switzerland.

World Health Organization, 2017. Health Topics: Diarrhoea. Available at: http://www.who.int/topics/diarrhoea/en/.

Zambrano, L.D. et al., 2014. Human diarrhea infections associated with domestic animal husbandry: A systematic review and meta-analysis. Transactions of the Royal Society of Tropical Medicine and Hygiene, 108(6), pp.313–325.

SECTION

6. CONCLUSION

6.1. DOCTORAL SUMMARY

The two goals of this dissertation were to 1) test the three primary hypotheses proposed below and 2) rank order the causal factors to child stunting. The three primary hypotheses tested in this dissertation included that among children in the western highlands of Guatemala between 0 and 5 years of age;

Hypothesis #1: there is a statistically significant association between the severity of the children's environmental enteric dysfunction (EED) and the ratio of the children's height-for-age.

Hypothesis #2: there is a statistically significant association between the children's aflatoxin B (AFB) exposure level and the ratio of the children's height-for-age.

Hypothesis #3: there is a statistically significant association between the children's AFB exposure level and the severity of the children's EED.

6.1.1 Goal One: The Primary Hypotheses.

Hypothesis #1

Based on the results of this study, the data partially confirmed Hypothesis #1, but further research is needed. Results reported in Paper I and part of Paper II were from an analysis in which a weighted correlation network algorithm was applied to data from a representative sample of five departments in the western highlands of Guatemala (Huehuetenango, San Marcos, Quiche, Totonicapán, and Quetzaltenango). Diarrheal occurrences and usage of oral rehydration therapy (ORT) were second level nodes to child height-for-age, suggesting they were important (no EED variable was available). When children were divided into levels of severity of stunting (not stunted = >-2 SD, stunted = -2 to -3 SD, severely stunted = <-3 SD), diarrheal occurrences and ORT use were second level nodes to the height-for-age of children that were not stunted. However, for stunted children only ORT use was important, while for severely stunted children, neither diarrheal occurrences nor ORT use was important. This suggests that the correlation between diarrheal occurrences and child height-for-age becomes less statistically detectable among a population of children who are all stunted or severely stunted. Finally, when children were divided by age in months (0-6, 7-12, 13-18, 19-24), the only age group with diarrheal occurrences within two levels of nodes of child height-for-age was children 0-6 months of age.

Results reported in Paper II and Paper III were from an analysis of a community called San Vicente in the western highlands of Guatemala. Reported results showed no statistically significant correlation between EED and child height-for-age. In Paper II a hypothesized SEM was tested to assess a cross-sectional model which included regressing child height-for-age on EED, AFB exposure, child diet diversity, prenatal health, and child play time. The model was tested with data collected in October 2016; EED was not statistically significant with child height-for-age. Additionally, in Paper III, a smaller SEM was tested which analyzed the hypothesized correlations between EED, AFB exposure, and child height-for-age both in a cross-sectional and two time-point methodology. Again, no relationship was found between EED and child height-for-age.

Finally, results reported in Paper V were from an analysis in which geospatially based water, sanitation, and hygiene (WaSH) infrastructure SEMs were built and tested with two datasets from a representative sample of five departments in the western highlands of Guatemala (as listed above). WaSH infrastructure variables significantly correlated with child height-for-age were hypothesized to affect the child by blocking enteric pathogens from reaching the child's intestines. When the five geographically based WaSH models focused on child height-for-age were built and tested with the two datasets, nine of the ten models showed at least one significant WaSH infrastructure variable as significant. This supports the hypothesis that EED may contribute to child stunting in some way; however, the magnitude and origin are still unknown.

Hypothesis #2

<u>Based on the results from this study, the data confirmed Hypothesis #2</u>. Results reported in Paper I that utilized the weighted correlation network algorithm on regional data showed that among groups of children 0-6 and 7-12 months of age, maize storage

was a first level node to child height-for-age. However, in groups above 12 months of age, maize storage was not closely related.

Results reported in Paper III were from bivariate Kruskal-Wallis significance tests, a cross-sectional SEM, and a two time-point SEM. The Kruskal-Wallis tests identified a significant correlation between the child height-for-age variable and a dichotomous variable created based on a set level for putative AFB consumption for the child (10 ng/kg of body weight). This correlation was significant for the child height-for-age in October 2016 and the set level for putative AFB consumption in October 2016, as well as the change in child height-for-age between October 2016 and February 2017 and the October 2016 set level for putative AFB consumption. Next, the hypothesized cross-sectional SEM was tested with data from October 2016 and identified that the computed continuous value of putative AFB consumption per weight of the child was negatively correlated with the child height-for-age (controlling for food consumption and socio-economic status). Furthermore, the two time-point SEM demonstrated similar results by identifying a significant negative correlation between the continuous value of putative AFB consumed per weight of child and the change in child height-for-age between October 2016 and February 2017. The results support a growing body of literature linking AFB exposure and child stunting.

Hypothesis #3

Based on the results from this study the data did not confirm Hypothesis #3. Results reported in Paper III from both the cross-sectional SEM and two time-point SEM demonstrated a negative correlation between the putative AFB consumption per weight of child and EED. Not only does this result not confirm Hypothesis #3, but is counter to it. The data suggest that for children who consume higher levels of AFB, the severity of their EED is lower. Results from Paper I and Paper II did not confirm Hypothesis #3 as well.

6.1.2 Goal Two: Rank Order of Causal Factors.

Child Stunting

Results from three different papers provide insight to prioritizing correlated variables hypothesized to affect a child’s height-for-age. First, for the region of the western highlands of Guatemala (five departments) results from Paper I provide insight among stunting severity levels and child age ranges. First, the three primary categories of variables that are different between stunted and non-stunted children are 1) food type and diversity, 2) farming practices and maize quality, and 3) intestinal health. Second, in addition to the categories above, when children are divided into age ranges (0-6 months, 7-12 months, etc.) the most consistent variables among all age ranges was the mother's height and weight.

Results reported in Paper II were from the analysis of a single community utilizing a hypothesized SEM for the child height-for-age model. The two variables most important to child height-for-age in San Vicente were the number of times the child played the day before (negative) and the prenatal health practices of the mother during pregnancy (positive). Specifically, these results suggest that 1) the sanitary conditions in which children play may be important to their long term physical development and 2) maintaining good prenatal health practices in that vitamins are taken and health checkups are attended may positively impact long term child physical development.

Finally, results reported in Paper V from the geographically based WaSH infrastructure models on child height-for-age provides insight to the infectious disease transmission pathways potentially significant to long term child physical development. From the final 2013 SEMs for the height-for-age models the data suggest floors and foods may be common transmission pathways across all geographies that affect the child's physical development.

Aflatoxin B Exposure

Results reported in Paper IV were from an analysis of San Vicente testing two hypothesized SEMs on the exposure pathways of AFB in maize to the child from 1) the local markets or 2) subsistence farming. The outcome variable used was a set of symptoms in common with high exposure to AFB. This variable was built and tested in Paper III by comparing putative AFB consumption levels among children to reported AFB symptoms. To test both the market and subsistence SEMs, two datasets for each time-point (October 2016 and February 2017) were created. In three of the four models (subsistence 2016, market 2016, and market 2017), the occurrence of the observing fungus by the field team in the household's maize storage was found to have a negative correlation with AFB symptoms among children. This may have been due to the fact that if households were also able to observe the fungus, they could avoid consuming it.

Results from the market SEM identified purchase habits of the mother at the market had the largest negative correlation with an increase in AFB symptoms among children. Furthermore, this was more important a month before harvest as compared to two months after harvest.

Finally, results from the subsistence SEMs identified several potentially contributing factors to AFB exposure. Post-harvest practices, including ideal maize drying time, improved maize drying surface, and maize shearing practices, had the largest negative effect size on increased AFB symptoms among children, while improve maize storage also contributed to a reduction in AFB symptoms. This suggested that in the region of San Vicente, both post-harvest practices and maize storage were important but that post-harvest practices may have had a more significant impact. Additionally, both storing maize to receive a higher price later and improving maize quality to receive a higher price at market had overall (total) negative effects on the AFB symptom variable. This supports the notion that if buyers at the market recognize the negative value of fungus in maize, the households who practice subsistence farming may have improved post-harvest practices or improved types of maize storage. Finally, these correlations were much stronger one month before harvest as compared to two months after harvest.

Increased EED Severity

Lastly, results from Paper II and Paper V support several areas within WaSH in which may contribute to EED in the western highlands of Guatemala. From Paper II, three primary categories were identified to be significant to a child in the western highlands of Guatemala having diarrhea. These categories include water availability, sanitation facilities, and pathogen transmission barriers. Furthermore, from Paper II the data from San Vicente confirmed the finding from the network analysis that the category of pathogen transmission barriers was important. This was confirmed by the identification in the cross-sectional analysis from October 2016, that utilizing improved types of water treatment had the largest statistically significant effect on reducing EED.

Results from Paper V supported the findings reported in Paper II in several areas. First, in 2012 over all regions the WaSH infrastructure variable, type of sanitation, was most commonly negatively correlated (five of ten models) to diarrhea (acute model) or EED (medium model). Similarly, the putative transmission pathway, floors, was most commonly important (six of ten scenarios) to diarrhea or EED. Additionally, San Vicente is located in a geographically similar region as San Marcos. For both the diarrhea and EED models for San Marcos in 2012, type of water treatment was negatively correlated with an increase in the harmful outcome variable (diarrhea or EED). These results suggest that a focus on water treatment among communities near San Vicente may be important for the intestinal health of local children while a focus on sanitation and floors may be important for both acute and chronic health in the larger region. Finally, these models offer a base platform in which to continue to improve a geographically unique understanding of potential causal factors to child health problems.

6.2. KEY TAKEAWAYS

The key takeaways from this dissertation include;

- The network analysis identified nutrition, maize farming practices, and diarrhea as trends related to growth
- Higher putative AFB exposure and AFB symptoms were negatively correlated with child height-for-age, no relationship was found between EED and child height-for-age
- The child height-for-age model identified prenatal health as beneficial and more child play times as negative for child height-for-age
- AFB models identified maize storage, post-harvest practices, and maize purchase habits as negatively correlated with increased AFB symptoms
- AFB transmission variables significant in ‘lean’ season, but not immediately following harvest; additionally, if fungus had a more recognized negative monetary value, household was less exposed
- EED models identified improved water treatment as negatively correlated with EED for the community of San Vicente
- Geospatial EED models identified clean floors and sanitation as most commonly negatively correlated with EED for the region of the western highlands

Practitioner Recommendations:

- Focus on prenatal health and healthy child play time trainings/programs for mothers
- Focus on economic growth among households as it can positively impact child health (captured as 'maize farming practices' in this study)
- Focus on 'floor' based infectious disease transmission pathways for western highlands (sanitation, animal pens, improved flooring, etc.) and water treatment for the San Vicente region
- Focus on public health advocacy for fungal toxins to help customers and farmers recognize a greater negative value in fungus (market based approach/intervention)

6.3. PROPOSED FUTURE WORK

Mycotoxins transmission/biomarker development in Guatemala – Beginning with a group of researchers including people from Duke University and Universidad de Rafael Landivar, the alignment of a AFB research agenda supported by NIH funding, USAID, and USDA will provide a unified research front on this emerging problem. Goal: set urinary AFB biomarker, conduct studies in three different research sites to demonstrate public health danger and the need for culturally sensitive public health action in Guatemala.

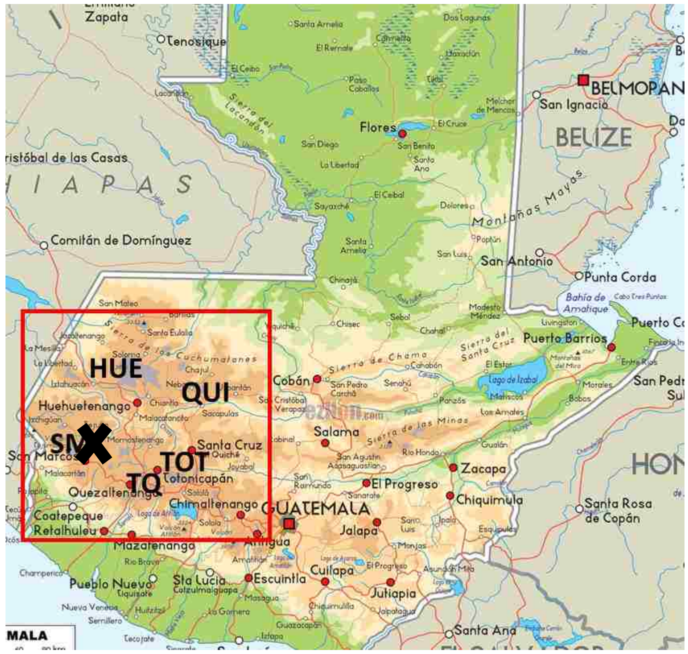
Environmental enteric dysfunction/intestinal health research in Guatemala – The health office at USAID Guatemala is interested in the continued support of research on this topic and coordinating several ongoing and upcoming projects throughout the western highlands. Goal: apply new research methods (including sensor combinations and modeling techniques) to elucidate the mechanistic pathways of cause and effect for EED in Guatemala.

EED/intestinal health and food security research in Southern Africa – University of Missouri, University of Western Cape, University of the Witwatersrand, and North-West University are developing a collaborative research agenda on the EED-nutrition interaction. The UN Scaling Up Nutrition Movement has interest and authorization to work in Southern Africa and has partnered with NWU in the past making them an excellent partner. The group of researchers from the above institutions will apply to the

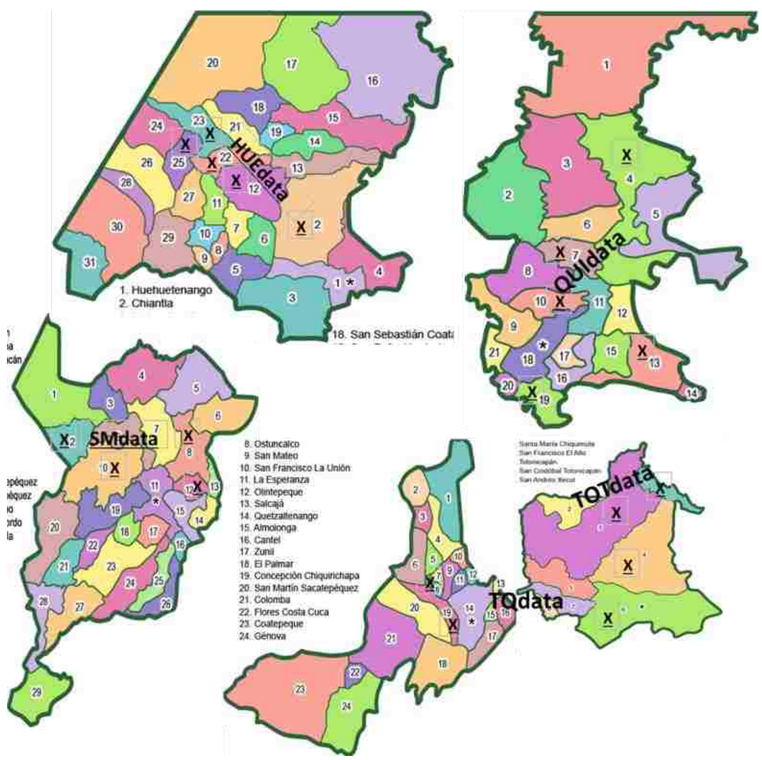
African Union/European Union Research Grant to seek funding for this project. Goal: duplicate Guatemala EED/intestinal health research in Southern Africa sites, incorporate more holistic nutrition analysis, and help coordinate regional partners (academic, NGO, and UN).

Microbiome/Fecal DNA sequencing research for Guatemala and Southern Africa – Arizona State University and the University of Missouri will begin collaborating on utilizing deep sequencing of DNA extract from fecal samples to analyze intestinal health of children in Guatemala. Goal: build out bioinformatics work for metagenomics, apply SEM approach to complex sequence data, and open new research sites to tackle potential biomarker development and mechanistic pathway investigations in Central America and Southern Africa.

APPENDIX A. MAP OF GUATEMALA



X indicates where San Vicente research site is located.



X's indicate municipalities in which the surveys were conducted for USAID in 2012 and 2013.

APPENDIX B. SURVEY INSTRUMENTS FROM SAN VICENTE HEALTH CENTER

Nombre de Madre: ____

Firma: ________________________ o Huella Dactilar: ___________

Años de Madre: ______

Nombre del niño #1: _______________________________________ Meses: _________ Sexo: **M** **F**

Nombre del niño #2: _______________________________________ (NA) Meses: _________ Sexo: **M** **F**

Nombre del niño #3: _______________________________________ (NA) Meses: _________ Sexo: **M** **F**

Otras personas que viven en su casa:

Abuela Si/No Años:_____

Abuelo Si/No Años:_____

Padre Si/No Años:_____

Cuantos otros personas _____ (niños>5 + adultos)

Tratamiento del Agua

¿Usted trata el agua de alguna manera para hacerla más segura para beber? Y SI ES ASÍ, CÓMO? (Marca con “X”)

___ cloro

___ hervirla

___ ponerla en el sol

___ filtro

___ reposarla en un recipiente

___ otro

¿Cuál es su fuente primaria de agua para tomar o beber (marca uno con "X")?

___ Pozo abierto

___ Pozo cerrado

___ Río

___ Botella/Jarrafon de agua

___ Sistema de distribución con chorro en la casa

___ Sistema de distribución con chorro afuera la casa

___ Otro

Vestimenta del bebé

¿Ayer, cuantas veces juega (sin distracciones; viajando, compras) con (NOMBRE)?

Niño #1: **0 1 2 3 4 5+** *Niño #2*: **0 1 2 3 4 5+ NA** *Niño #3*: **0 1 2 3 4 5+ NA**

¿Cuál es la forma correcta que usa para "cargar" a un niño menor de un año(a)? (Marca uno con "X")

___ Solo con el cargador

___ Con el cargador envuelto en una frazada floja – algo de movimiento

___ Con el cargador envuelto en una frazada un poco apretada – movimiento limitado

___ Con el cargador envuelto en una frazada apretado – sin movimiento

Amamantamiento/ Alimentación complementaria

¿A que edad alimenta a su niño con leche de formula u otro tipo de líquido?

Niño #1: ________________ *Niño #2*: ________________ (NA) *Niño #3*: ________________ (NA)

¿A los cuantos meses de edad, su niño dejo de tomar leche materna?

Niño #1: ________________ *Niño #2*: ________________ (NA)

Niño #3: ________________ (NA)

Salud durante el embarazo

¿Cuántas visitas de control prenatal tuvo usted durante el embarazo de (NOMBRE)?

Niño #1: ________________ *Niño #2*: ________________ (NA)

Niño #3: ________________ (NA)

¿Cuándo estaba embarazada de (NOMBRE), tomó usted (marca con “X”):

___ sulfato ferroso/hierro?

___ ácido fólico?

___ pastillas prenatales?

___ otro medicinas?

___ Nada

¿En Dónde tuvo lugar fue el parto de (NOMBRE) (marca una con “X”):

___ Casa

___ Hospital/Centro de Salud

___ ¿Otro

¿Cuánto tiempo después de nacido (NOMBRE) empezó a darle el pecho?

____

Toma calórica, micronutriente, proteínica

¿Fue ayer un día común, normal? **Si o No**

Ayer durante el día y la noche, consumió (NOMBRE) algún… **Si = 1; No = 0**

N#1 N#2 N#3

¿Alimentos hechos de granos tales como tortillas, tamalitos pan, arroz, fideos, cereales? ___ ___ ___

¿Papa, yuca, ichintal, camote blanco, otras raíces/tubérculos o alimentos hechos de raíces o tubérculos? ___ ___ ___

¿Vegetables?

___ ___ ___

¿Frutas?

___ ___ ___

¿Carne, tal como aves, res, chivo, cerdo, conejo?

___ ___ ___

¿Huevos?

___ ___ ___

¿Pescado fresco o seco, mariscos o comida de mar?

___ ___ ___

¿Alimentos hechos de frijoles, manías, lentejas, habas, arvejas, nueces o semillas?

___ ___ ___

¿Queso, crema, leche de vaca (liquida o en polvo), leche de cabra, yogurt u otros productos lácteos? ___ ___ ___

¿Aceite, mantequilla, margarina, manteca o alimentos hechos con cualquiera de estos productos? ___ ___ ___

¿Alimentos azucarados tales como chocolates, dulces, caramelos, pasteles, tortas o bizcochos? ___ ___ ___

¿Condimentos para sabor tales como chile, condimentos, hierbas aromáticas, polvo de pescado? ___ ___ ___

¿Cuántas tortillas consumió NINO ayer?

___ ___ ___

¿Cuántos tamalitos consumió NINO ayer?

___ ___ ___

En los últimos 30 días, ¿en algún momento no hubo comida de ningún tipo en su casa debido a la falta de recursos para conseguirla?

Sí **No**

Salud infantil

¿Ha tenido (NOMBRE) asientos o diarrea en los últimos dos días?

Niño #1: **Si o No** *Niño #2*: **Si o No (NA)** *Niño #3*: **Si o No (NA)**

¿Cuántos veces (NOMBRE) asientos o diarrea en las últimas dos semanas?

N#1:_________ *N#2*:________ *N#3*:________

¿(NOMBRE) tuvo doler inexplicable en su estómago o intestinos en las últimas dos semanas?

Niño #1: **Si o No** *Niño #2*: **Si o No (NA)** *Niño #3*: **Si o No (NA)**

¿Cuántos veces (NOMBRE) asientos o diarrea con sangre en el ultima mes?

N#1:_________ *N#2*:________ *N#3*:________

¿De qué se enferman más seguido los miembros de la familia? **gripe, tos, diarrea, neumonía, otras ___________**

¿Ha estado (NOMBRE) enfermo con fiebre en la última semana?

Niño #1: **Si o No** *Niño #2*: **Si o No (NA)** *Niño #3*: **Si o No (NA)**

¿Ha estado (NOMBRE) enfermo con tos en la última semana?

Niño #1: **Si o No** *Niño #2*: **Si o No (NA)** *Niño #3*: **Si o No (NA)**

Exposición a Hongos

¿Alguien en la casa menor de cincuenta años ha tenido problemas con (marca con “X”):

	***Niño #1*:**	***Niño #2*: (NA)**	***Niño #3*: (NA)**	***Otros Adultos:***
Orinar?	____	____	____	____
Dolores de cabeza crónicos?	____	____	____	____
Perdida de peso o apetito no intencional?	____	____	____	____

Los ojos o la piel se ha puesto amarilla?	____	____	____	____
Inexplicable picazón excesiva?	____	____	____	____

Nivel educativo de la madre

¿Cuál es el último año de estudios que usted ganó? (Marca con un círculo)

Nunca escuela Primario - 1 2 3 4 5 6 Básico - 1 2 3 Diversificado - 4 5 6 o completo

Maíz

¿De las dos semanas anteriores, el maíz que consumió es; **propio comprado**

¿Durante secando, por cuanto tiempo ustedes secan la mayoría del maíz? ________________

¿Durante secando, que tipo de superficie ustedes usan? **Tierra Techo Tablas Piso de Cemento Lona Otra**

¿Qué forma ustedes usan para desgranar el maíz? **Aporreo Desgranando Maquina Otro**

¿Ustedes recibieron un precio mejor para del maíz cuando lo almacenado? **Si o No**

¿Por cuántos meses (promedio) lo deja almacenado / guardado? _____ meses

¿Las tortillas que consumió ayer, fue; **comprado maseca**?

¿Ustedes recibieron más dinero por mejorar calidad maíz? **Si o No**

¿Tiene problemas con hongos en su maíz? **Si o No**

¿Cuáles son las 3 cosas en las que usted piensa/busca cuando compra/usa maíz? Enumérelos por orden de importancia.

1. **Hongos/Podrido Insectos/Animales Seco Limpieza/Calidad Precio Tamaño Otro**

2. **Hongos/Podrido** **Insectos/Animales** **Seco** **Limpieza/Calidad** **Precio** **Tamaño** **Otro**

3. **Hongos/Podrido** **Insectos/Animales** **Seco** **Limpieza/Calidad** **Precio** **Tamaño** **Otro**

Maíz del Mercado

¿De qué Mercado consigue su maíz? **Paloma** **Pologua** **Centro** **Sija** **Propio Maíz**

¿Hay hongos/moho visibles en el maíz del mercado? **Si o No**

Medidas Antropomórficas del niño(a) y la Madre

Altura del: *niño #1* ___________ cm *niño #2* ___________ cm *niño #3* ___________ cm la Madre __________ cm

Acostado / Derecho **Acostado / Derecho** **Acostado / Derecho**

Peso del: *niño #1* ___________ cm *niño #2* ___________ cm *niño #3* ___________ cm la Madre __________ cm

Muchas gracias por la participación del usuario.

Cuestionario de Observación (Visitas de Casa)

1. Aseo

¿Qué tipo de servicio sanitario tiene en su casa?

Inodoro, **Letrina con Piso de Cemento,** **Letrina con Piso de Madera,** **Sin Letrina**

¿A cuántos metros queda su el baño de;

La casa? **Menos de 10 metros** **Más de 10 metros** **50 a 100 metros** **Más 100 metros**

La fuente del agua? **Menos de 10 metros** **Más de 10 metros** **50 a 100 metros** **Más 100 metros**

2. Prácticas de higiene

¿Hay jabón de manos en donde se lava las manos? **Si o No**

¿Hay agua disponible en dónde se lava las manos? **Si o No**

¿En el suelo de la casa se observa;

Excremento de Animal? **Si o No**

Animales? **Si o No**

Basura? **Si o No**

Moscas? **Si o No**

¿A qué distancia está el lugar donde se lava las manos de;

El sanitario? **Menos de 10 metros** **Más de 10 metros** **50 a 100 metros** **Más 100 metros**

La cocina? **Menos de 10 metros** **Más de 10 metros** **50 a 100 metros** **Más 100 metros**

3. Fuente de Agua

¿Tiene un tanque para agua? **Si o No**

¿A cuántos metros queda su fuente de agua de;

La casa? **Menos de 10 metros** **Más de 10 metros** **50 a 100 metros** **Más 100 metros**

El corral? **Menos de 10 metros** **Más de 10 metros** **50 a 100 metros** **Más 100 metros**

4. Almacenamiento de Agua

¿Qué clase de recipientes usa usted para almacenar el agua?

Plástico con tapa **Plástico sin tapa** **Barro con tapa** **Barro sin tapa** **Metal con tapa** **Metal sin tapa**

¿Cómo saca el agua de los recipientes? **Grifo o chorro, Cucharon, Taza, verter el agua, otro**

5. Preparación de la comida

¿Tiene en su casa un lugar (cuarto) que usan solo para cocinar? **Si o No**

¿Hay alguno de los siguientes en su cocina;

Animales? **Si o No**

Moscas? **Si o No**

Basura? **Si o No**

Suelo sucio? **Si o No**

¿Qué tipo de depósito usa para maíz? **Cajón, sacos, toneles, silo, otros**

¿Desde la última cosecha ha visto hongos en su maíz? **Si o No**

6. Estatus socio-económico

¿Qué tipo de piso tiene la vivienda? **Tierra, torta de cemento, piso cerámico, madera, otro**

¿Qué tipo de paredes tiene la vivienda? **No tiene paredes, barro, madera, caña, blok/concreto, ladrillo, otro**

¿Qué tipo de techo tiene la vivienda? **Palma, lamina, teja de barro, concreto, otros**

¿Cuántos cuartos (habitaciones) usan para dormir? **1 2 3 4 5 6+**

¿Tiene la vivienda una televisión? **Si o No**

¿Cuántas bombillas/luces tiene la vivienda? **1 2 3 4 5 6 7 8+**

APPENDIX C. SENSITIVITY ANALYSIS OF SEMS

A random subsample of 75% of the original data was taken for the individual datasets used to assess each model. For each SEM, 20 iterations were conducted and outputs were recorded including the model fit tests, parameter estimates, and p-values. For each correlation in the model a mean, minimum, maximum, standard deviation, and confidence interval were computed. Excel 2016 was used for the exercise.

Child Height-for-Age SEM presented in Paper II

Sensitivity Analysis of Stunting SEM 2016								*did not converge					
								^reason for nonconvergence					
	Model Fit Results (Robust score				**HAZ on Child Play**			**HAZ on Prenatal Health**			**HAZ on AFB Burden**		
	Chi-Squa	**RMSEA**	**CFI**	**TLI**	**Unst. Est.**	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**
Original - Full Data					-0.085	-0.092	0.076	-1.629	-0.151	0.102	-0.333	-0.02	0.845
VG16STDataSUB1*	-	-	-	-	-0.053	-0.058	-	-3.326	-0.219	-	^	^	^
VG16STDataSUB2	0.297	0.017 (.(	0.971	0.961	-0.075	-0.077	0.208	-1.915	-0.178	0.055	-1.724	-0.076	0.622
VG16STDataSUB3	0.312	0.016 (.(	0.975	0.967	-0.052	-0.055	0.346	-1.343	-0.102	0.33	0.176	0.013	0.899
VG16STDataSUB4	0.043	0.032 (.(	0.912	0.882	-0.117	-0.123	0.041	-1.864	-0.16	0.168	-0.376	-0.025	0.847
VG16STDataSUB5*	-	-	-	-	-0.087	-0.097	-	-3.353	-0.262	-	^	^	^
VG16STDataSUB6*	-	-	-	-	-0.122	-0.128	-	-1.689	-0.175	-	^	^	^
VG16STDataSUB7	0.033	0.034 (.(	0.904	0.87	-0.097	-0.107	0.079	-2.315	-0.214	0.028	-0.516	-0.039	0.751
VG16STDataSUB8	0.046	0.032 (0	0.903	0.869	-0.14	-0.152	0.014	-1.459	-0.147	0.125	-1.975	-0.114	0.361
VG16STDataSUB9*	-	-	-	-	-0.107	-0.121	-	-1.342	-0.14	-	^	^	^
VG16STDataSUB10	0.005	0.041 (0	0.834	0.776	-0.058	-0.061	0.316	-1.818	-0.159	0.146	-2.509	-0.11	0.435
VG16STDataSUB11*	-	-	-	-	-0.064	-0.068	-	-2.955	-0.262	-	^	^	^
VG16STDataSUB12	0.077	0.029 (0	0.918	0.889	-0.098	-0.106	0.061	-1.273	-0.151	0.154	-2.183	-0.125	0.371
VG16STDataSUB13	0.195	0.022 (0	0.955	0.939	-0.121	-0.134	0.025	-1.691	-0.162	0.125	0.005	0	0.997
VG16STDataSUB14	0.525	0.000 (0	1	1.01	-0.077	-0.081	0.152	-2.175	-0.221	0.026	-1.887	-0.094	0.533
VG16STDataSUB15	0.209	0.021 (0	0.953	0.936	-0.095	-0.105	0.087	-0.704	-0.081	0.435	1.899	0.08	0.426
VG16STDataSUB16	0.027	0.034 (0	0.888	0.848	-0.078	-0.085	0.159	-2.384	-0.2	0.096	1.923	0.084	0.565
VG16STDataSUB17	0.197	0.022 (0	0.959	0.944	-0.075	-0.082	0.193	-1.277	-0.135	0.181	-0.256	-0.017	0.886
VG16STDataSUB18	0.089	0.028 (0	0.922	0.894	-0.081	-0.089	0.141	-1.448	-0.122	0.275	-0.106	-0.004	0.971
VG16STDataSUB19	0.52	0.000 (0	1	1.01	-0.089	-0.098	0.117	-2.366	-0.179	0.055	-0.186	-0.009	0.936
VG16STDataSUB20	0.009	0.039 (0	0.856	0.805	-0.104	-0.112	0.068	-4.244	-0.235	0.094	10.196*	-	0.233
Mean	0.17227		0.93	0.907	-0.0895	-0.09695	0.1338	-2.04705	-0.1752	0.152867	-0.55107	-0.03114	0.655533
Min	0.005		0.834	0.776	-0.14	-0.152	0.014	-4.244	-0.262	0.026	-2.509	-0.125	0.233
Max	0.525		1	1.01	-0.052	-0.055	0.346	-0.704	-0.081	0.435	1.923	0.084	0.997
Standard Deviation	0.17458		0.0492	0.069	0.024278	0.026743	0.099341	0.868979	0.049614	0.115281	1.389431	0.066744	0.257022
Confid. Interval	0.07651		0.0216	0.03	0.01064	0.011721	0.043537	0.38084	0.021744	0.050523	0.608934	0.029251	0.112643
Upper CI	0.24878		0.9516	0.937	-0.07886	-0.08523	0.177337	-1.66621	-0.15346	0.20339	0.057863	-0.00189	0.768176
Lower CI	0.09575		0.9084	0.877	-0.10014	-0.10867	0.090263	-2.42789	-0.19694	0.102343	-1.16001	-0.06039	0.542891

Child Height-for-Age SEM presented in Paper II (cont.)

HAZ on EED			**HAZ on Diet Diversity**			**EED on Prenatal Health**			**EED on AFB Burden**		
Unst. Est.	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**
0.02	0.005	0.918	-0.059	-0.047	0.312	0.264	0.096	0.206	-0.202	-0.048	0.459
0.037	0.011	-	-0.032	-0.025	-	0.231	0.05	-	-	-	-
0.04	0.009	0.868	-0.031	-0.024	0.647	0.231	0.097	0.262	-0.537	-0.107	0.202
-0.16	-0.038	0.547	-0.048	-0.038	0.474	0.435	0.139	0.166	-0.177	-0.055	0.474
-0.109	-0.027	0.651	-0.084	-0.068	0.169	0.419	0.147	0.117	-0.302	-0.081	0.271
-0.105	-0.034	-	-0.078	0.065	-	0.04	0.01	-	-	-	-
0.084	0.023	-	-0.048	-0.037	-	0.335	0.127	-	-	-	-
-0.075	-0.017	0.753	-0.017	-0.014	0.809	0.22	0.088	0.291	-0.197	-0.064	0.338
-0.141	-0.037	0.569	-0.027	-0.022	0.663	0.521	0.201	0.046	-0.365	-0.08	0.215
0.073	0.025	-	0.028	0.022	-	0.279	0.085	-	-	-	-
0.07	0.017	0.782	-0.04	-0.03	0.554	0.149	0.054	0.546	-0.458	-0.083	0.297
0.135	0.038	-	-0.085	-0.065	-	0.184	0.058	-	-	-	-
-0.006	-0.001	0.981	-0.066	-0.052	0.33	0.184	0.098	0.221	-0.419	-0.108	0.2
0.105	0.026	0.619	-0.044	-0.035	0.501	0.31	0.118	0.182	-0.174	-0.053	0.452
-0.047	-0.011	0.854	-0.107	-0.083	0.124	0.236	0.101	0.249	-0.572	-0.119	0.272
-0.084	-0.02	0.724	-0.093	-0.074	0.17	0.217	0.105	0.177	-0.613	-0.109	0.182
0.098	0.023	0.675	-0.073	-0.058	0.256	0.289	0.102	0.294	-0.669	-0.123	0.176
0.173	0.05	0.396	-0.09	-0.071	0.198	0.42	0.152	0.073	-0.237	-0.054	0.526
0.09	0.021	0.708	-0.123	-0.092	0.087	0.475	0.17	0.109	-0.785	-0.135	0.173
0.029	0.007	0.914	-0.049	-0.039	0.483	0.229	0.07	0.417	-0.472	-0.091	0.257
0.315	0.083	0.217	-0.058	-0.047	0.365	-0.196	-0.041	0.634	1.930*	-	0.328
0.0261	0.0074	0.683867	-0.05825	-0.03935	0.388667	0.2604	0.09655	0.252267	-0.42693	-0.09014	0.290867
-0.16	-0.038	0.217	-0.123	-0.092	0.087	-0.196	-0.041	0.046	-0.785	-0.135	0.173
0.315	0.083	0.981	0.028	0.065	0.809	0.521	0.201	0.634	-0.174	-0.053	0.526
0.11852	0.031277	0.199791	0.034952	0.036329	0.221985	0.160927	0.055262	0.167973	0.194635	0.027352	0.113758
0.051943	0.013708	0.08756	0.015318	0.015922	0.097288	0.070528	0.024219	0.073616	0.085301	0.011987	0.049856
0.078043	0.021108	0.771427	-0.04293	-0.02343	0.485954	0.330928	0.120769	0.325883	-0.34163	-0.07816	0.340723
-0.02584	-0.00631	0.596306	-0.07357	-0.05527	0.291379	0.189872	0.072331	0.17865	-0.51223	-0.10213	0.241011

EED SEM presented in Paper II

Sensitivity Analysis of EED SEM 2016									
	Model Fit Results (Robust scores)				EED on WaterTreat			EED on WaterSource	
	Chi-Squar	RMSEA	CFI	TLI	Unst. Est.	Stnd. Est.	P-value	Unst. Est.	Stnd. Est.	P-value
Original - Full Data	0.092	0.026 (0.0(	0.986	0.975	0.09	0.113	0.01	-0.068	-0.107	0.079
VG16EEDsub1	0.211	0.023 (0.0(	0.99	0.982	0.084	0.106	0.038	-0.055	-0.082	0.239
VG16EEDsub2	0.252	0.020 (0.0(	0.993	0.988	0.082	0.115	0.044	-0.062	-0.109	0.171
VG16EEDsub3	0.302	0.017 (0.0(	0.994	0.989	0.084	0.116	0.029	-0.047	-0.071	0.338
VG16EEDsub4	0.676	0.000 (0.0(	1	1.013	0.097	0.1	0.03	-0.096	-0.142	0.043
VG16EEDsub5	0.055	0.035 (0.0(	0.972	0.95	0.087	0.125	0.001	-0.026	-0.047	0.533
VG16EEDsub6	0.115	0.029 (0.0(	0.981	0.966	0.079	0.097	0.079	-0.062	-0.097	0.17
VG16EEDsub7	0.051	0.035 (0.0(	0.973	0.953	0.07	0.08	0.056	-0.095	-0.145	0.036
VG16EEDsub8	0.099	0.030 (0.0(	0.981	0.966	0.075	0.08	0.124	-0.063	-0.096	0.211
VG16EEDsub9	0.238	0.021 (0.0(	0.99	0.983	0.071	0.092	0.135	-0.032	-0.059	0.435
VG16EEDsub10	0.098	0.030 (0.0(	0.98	0.966	0.097	0.127	0.025	-0.038	-0.059	0.432
VG16EEDsub11	0.386	0.012 (0.0(	0.997	0.995	0.058	0.077	0.151	-0.06	-0.104	0.18
VG16EEDsub12	0.205	0.024 (0.0(	0.987	0.977	0.094	0.111	0.025	-0.037	-0.055	0.441
VG16EEDsub13	0.142	0.027 (0.0(	0.985	0.974	0.081	0.106	0.055	-0.038	-0.053	0.488
VG16EEDsub14	0.319	0.017 (0.0(	0.993	0.988	0.128	0.154	0.001	-0.105	-0.174	0.005
VG16EEDsub15	0.162	0.026 (0.0(	0.986	0.976	0.108	0.135	0.018	-0.07	-0.106	0.198
VG16EEDsub16	0.478	0.000 (0.0(	1	1.001	0.097	0.132	0.021	-0.068	-0.104	0.126
VG16EEDsub17	0.236	0.021 (0.0(	0.991	0.984	0.059	0.083	0.018	-0.094	-0.141	0.025
VG16EEDsub18	0.259	0.020 (0.0(	0.993	0.987	0.151	0.177	0	-0.11	0.168	0.014
VG16EEDsub19	0.509	0.000 (0.0(	1	1.003	0.079	0.114	0.032	-0.048	-0.084	0.22
VG16EEDsub20	0.107	0.030 (0.0(	0.982	0.968	0.093	0.106	0.033	-0.104	-0.134	0.054
Mean	0.245		0.9884	0.98045	0.0887	0.11165	0.04575	-0.0655	-0.0847	0.21795
Min	0.051		0.972	0.95	0.058	0.077	0	-0.11	-0.174	0.005
Max	0.676		1	1.013	0.151	0.177	0.151	-0.026	0.168	0.533
Standard Deviation	0.16409		0.008369	0.016204	0.021844	0.025475	0.04384	0.026645	0.069299	0.171577
Confid. Interval	0.071914		0.003668	0.007102	0.009573	0.011165	0.019214	0.011677	0.030371	0.075195
Upper CI	0.316914		0.992068	0.987552	0.098273	0.122815	0.064964	-0.05382	-0.05433	0.293145
Lower CI	0.173086		0.984732	0.973348	0.079127	0.100485	0.026536	-0.07718	-0.11507	0.142755

EED SEM presented in Paper II (cont.)

EED on Food Prep			**FoodPrep on WaterStorage**			**FoodPrep on WaterSource**			**WaterTreat on WaterSource**		
Unst. Est.	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**
-0.13	-0.119	0.171	-0.103	-0.516	0	-0.235	-0.405	0	0.051	0.064	0.019
-0.195	-0.162	0.119	-0.093	-0.507	0	-0.225	-0.402	0	0.029	0.034	0.147
-0.131	-0.138	0.189	-0.109	-0.531	0	-0.291	-0.484	0	0.07	0.087	0.017
-0.079	-0.076	0.482	-0.114	-0.565	0	-0.266	-0.42	0	0.038	0.041	0.078
-0.082	-0.067	0.514	-0.099	-0.518	0	-0.227	-0.407	0	0.062	0.089	0.017
-0.071	-0.069	0.507	-0.09	-0.481	0	-0.246	-0.451	0	0.069	0.085	0.018
-0.117	-0.1	0.322	-0.102	-0.517	0	-0.234	-0.42	0	0.049	0.061	0.078
-0.113	-0.102	0.329	-0.116	-0.523	0	-0.269	-0.452	0	0.065	0.087	0.01
-0.13	-0.115	0.244	-0.091	-0.439	0	-0.246	-0.42	0	0.034	0.049	0.069
-0.107	-0.113	0.28	-0.108	-0.519	0	-0.261	-0.459	0	0.03	0.043	0.099
-0.116	-0.105	0.311	-0.106	-0.534	0	-0.24	-0.405	0	0.054	0.063	0.073
-0.124	-0.112	0.299	-0.089	-0.494	0	-0.25	-0.477	0	0.064	0.082	0.025
-0.017	-0.015	0.889	-0.103	-0.521	0	-0.237	-0.405	0	0.023	0.029	0.266
-0.131	-0.12	0.236	-0.102	-0.526	0	-0.258	-0.399	0	0.059	0.064	0.072
-0.193	-0.195	0.037	-0.094	-0.444	0	-0.207	-0.34	0	0.056	0.077	0.037
-0.098	-0.088	0.418	-0.113	-0.534	0	-0.283	-0.475	0	0.056	0.068	0.049
-0.034	-0.028	0.801	-0.095	-0.515	0	-0.2	-0.374	0	0.053	0.06	0.072
-0.132	-0.123	0.213	-0.103	-0.516	0	-0.217	-0.349	0	0.07	0.074	0.029
-0.193	-0.174	0.087	-0.116	-0.555	0	-0.24	-0.407	0	0.046	0.06	0.024
-0.158	-0.165	0.108	-0.098	-0.489	0	-0.231	-0.393	0	0.02	0.025	0.313
-0.264	-0.194	0.049	-0.093	-0.496	0	-0.217	-0.381	0	0.055	0.063	0.075
-0.12425	-0.11305	0.3217	-0.1017	-0.5112	0	-0.24225	-0.416	0	0.0501	0.06205	0.0784
-0.264	-0.195	0.037	-0.116	-0.565	0	-0.291	-0.484	0	0.02	0.025	0.01
-0.017	-0.015	0.889	-0.089	-0.439	0	-0.2	-0.34	0	0.07	0.089	0.313
0.05752	0.049268	0.229454	0.008862	0.03125	0	0.024137	0.040338	0	0.015921	0.019922	0.08022
0.025209	0.021592	0.100561	0.003884	0.013696	#NUM!	0.010579	0.017679	#NUM!	0.006977	0.008731	0.035157
-0.09904	-0.09146	0.422261	-0.09782	-0.4975	#NUM!	-0.23167	-0.39832	#NUM!	0.057077	0.070781	0.113557
-0.14946	-0.13464	0.221139	-0.10558	-0.5249	#NUM!	-0.25283	-0.43368	#NUM!	0.043123	0.053319	0.043243

Three-way SEM of HAZ, EED, and AFB exposure in Paper III

Sensitivity Analysis of HAZ-EED-AFB SEM 2016													
	Model Fit Results (Robust scores)				**HAZ on EED**			**HAZ on AFB**			**EED on AFB**		
	Chi-Squar	**RMSEA**	**CFI**	**TLI**	**Unst. Est.**	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**
Original - Full Data	0.21	0.034 (0.0(	0.966	0.916	0.297	0.036	0.588	-0.765	-0.076	0.032	-0.086	-0.071	0.051
VG16hypothSUB1	0.568	0.000 (0.0(	1	1.061	0.293	0.041	0.577	-0.415	-0.048	0.211	-0.114	-0.094	0.03
VG16hypothSUB2	0.336	0.021 (0.0(	0.986	0.966	-0.031	-0.003	0.957	-0.78	-0.085	0.03	-0.069	-0.067	0.064
VG16hypothSUB3	0.739	0.000 (0.0(	1	1.112	0.026	0.003	0.962	-0.628	-0.07	0.089	-0.093	-0.082	0.028
VG16hypothSUB4	0.217	0.035 (0.0(	0.968	0.919	0.058	0.009	0.88	-0.808	-0.089	0.018	-0.091	-0.062	0.075
VG16hypothSUB5	0.122	0.050 (0.0(	0.908	0.771	0.303	0.04	0.621	-0.857	-0.097	0.024	-0.057	-0.049	0.251
VG16hypothSUB6	0.829	0.000 (0.0(	1	1.317	0.977	0.094	0.297	-0.412	-0.043	0.215	-0.071	-0.077	0.114
VG16hypothSUB7	0.051	0.070 (0.0(	0.863	0.659	0.128	0.019	0.79	-0.935	-0.067	0.242	-0.169	-0.08	0.097
VG16hypothSUB8	0.32	0.021 (0.0(	0.988	0.969	0.869	0.095	0.208	-0.8	-0.085	0.023	-0.09	-0.087	0.047
VG16hypothSUB9	0.381	0.011 (0.0(	0.996	0.991	0.267	0.031	0.674	-0.884	-0.093	0.013	-0.094	-0.084	0.03
VG16hypothSUB10	0.144	0.050 (0.0(	0.938	0.845	0.206	0.025	0.723	-1.351	-0.086	0.14	-0.156	-0.081	0.084
VG16hypothSUB11	0.586	0.000 (0.0(	1	1.075	0.297	0.035	0.59	-0.639	-0.07	0.094	-0.11	-0.102	0.037
VG16hypothSUB12	0.215	0.038 (0.0(	0.954	0.884	-0.277	-0.035	0.571	-0.769	-0.086	0.034	-0.094	-0.084	0.029
VG16hypothSUB13	0.659	0.000 (0.0(	1	1.076	0.964	0.096	0.198	-1.069	-0.103	0.011	-0.089	-0.086	0.071
VG16hypothSUB14	0.526	0.000 (0.0(	1	1.034	0.69	0.084	0.276	-0.946	-0.098	0.02	-0.095	-0.081	0.044
VG16hypothSUB15	0.244	0.033 (0.0(	0.962	0.905	0.105	0.01	0.899	-0.65	-0.072	0.07	-0.068	-0.08	0.065
VG16hypothSUB16	0.185	0.040 (0.0(	0.961	0.901	0.501	0.071	0.358	-0.631	-0.073	0.079	-0.108	-0.088	0.033
VG16hypothSUB17	0.116	0.056 (0.0(	0.916	0.791	0.537	0.063	0.416	-0.869	-0.059	0.351	-0.102	-0.059	0.243
VG16hypothSUB18	0.546	0.000 (0.0(	1	1.053	0.114	0.014	0.835	-0.953	-0.068	0.269	-0.141	-0.08	0.074
VG16hypothSUB19	0.346	0.018 (0.0(	0.99	0.975	0.861	0.105	0.191	-0.657	0.069	0.091	-0.063	-0.054	0.186
VG16hypothSUB20	0.324	0.022 (0.0(	0.986	0.964	0.552	0.067	0.394	-0.919	-0.096	0.022	-0.089	-0.076	0.058
Mean	0.3727		0.9708	0.9634	0.372	0.0432	0.57085	-0.7986	-0.07095	0.1023	-0.09815	-0.07765	0.083
Min	0.051		0.863	0.659	-0.277	-0.035	0.191	-1.351	-0.103	0.011	-0.169	-0.102	0.028
Max	0.829		1	1.317	0.977	0.105	0.962	-0.412	0.069	0.351	-0.057	-0.049	0.251
Standard Deviation	0.224106		0.038277	0.142297	0.357957	0.03934	0.265303	0.217075	0.036845	0.101568	0.029457	0.013283	0.067349
Confid. Interval	0.098217		0.016775	0.062363	0.156879	0.017241	0.116272	0.095136	0.016148	0.044513	0.01291	0.005822	0.029516
Upper CI	0.470917		0.987575	1.025763	0.528879	0.060441	0.687122	-0.70346	-0.0548	0.146813	-0.08524	-0.07183	0.112516
Lower CI	0.274483		0.954025	0.901037	0.215121	0.025959	0.454578	-0.89374	-0.0871	0.057787	-0.11106	-0.08347	0.053484

AFB symptoms from Subsistence Farming SEM in Paper IV

Sensitivity Analysis of HAZ-EED-AFB SEM 2016										
	Model Fit Results (Robust scores)				AFB on ROI			AFB on ViewFungus		
	Chi-Squar	RMSEA	CFI	TLI	Unst. Est.	Stnd. Est.	P-value	Unst. Est.	Stnd. Est.	P-value
Original - Full Data	0.097	0.029 (0.00	0.954	0.92	0.021	0.07	0.039	-0.036	-0.11	0.004
VG16AFBfarmSUB1	0.037	0.042 (0.00	0.899	0.824	0.018	0.062	0.144	-0.032	-0.107	0.014
VG16AFBfarmSUB2	0	0.066 (0.04	0.763	0.587	0.012	0.037	0.495	-0.038	-0.107	0.015
VG16AFBfarmSUB3	0.038	0.040 (0.00	0.902	0.83	0.016	0.057	0.148	-0.037	-0.117	0.012
VG16AFBfarmSUB4	0.091	0.034 (0.00	0.937	0.891	0.02	0.06	0.091	-0.036	-0.091	0.012
VG16AFBfarmSUB5	0.017	0.044 (0.01	0.891	0.81	0.032	0.091	0.023	-0.042	-0.111	0.007
VG16AFBfarmSUB6	0.175	0.027 (0.00	0.964	0.938	0.022	0.074	0.084	-0.043	-0.132	0.005
VG16AFBfarmSUB7	0.004	0.054 (0.00	0.857	0.75	0.004	0.013	0.775	-0.049	-0.134	0.009
VG16AFBfarmSUB8	0.335	0.017 (0.00	0.984	0.971	0.016	0.044	0.129	-0.036	-0.097	0.013
VG16AFBfarmSUB9	0.195	0.026 (0.00	0.963	0.936	0.025	0.07	0.073	-0.034	-0.09	0.01
VG16AFBfarmSUB10	0.057	0.038 (0.00	0.923	0.865	0.022	-	0.061	-0.043	-	0.006
VG16AFBfarmSUB11*	-	-	-	-						
VG16AFBfarmSUB12	0.181	0.027 (0.00	0.958	0.928	0.018	0.051	0.16	-0.043	-0.11	0.007
VG16AFBfarmSUB13	0.132	0.031 (0.00	0.937	0.891	0.02	-	0.131	-0.041	-	0.006
VG16AFBfarmSUB14	0.008	0.047 (0.02	0.894	0.816	0.017	-	0.144	-0.038	-	0.009
VG16AFBfarmSUB15	0.156	0.029 (0.00	0.95	0.913	0.018	0.059	0.15	-0.034	-0.107	0.017
VG16AFBfarmSUB16	0.242	0.028 (0.00	0.97	0.948	0.025	0.072	0.075	-0.044	-0.116	0.007
VG16AFBfarmSUB17	0.007	0.052 (0.02	0.849	0.737	0.014	0.036	0.3	-0.046	-0.112	0.009
VG16AFBfarmSUB18	0.004	0.054 (0.03	0.824	0.694	0.004	0.01	0.826	-0.055	-0.123	0.011
VG16AFBfarmSUB19	0.416	0.010 (0.00	0.993	0.988	0.018	0.072	0.115	-0.042	-0.15	0.013
VG16AFBfarmSUB20	0.226	0.024 (0.00	0.965	0.94	0.015	0.047	0.214	-0.038	-0.111	0.012
Mean	0.122158		0.917	0.855632	0.017684	0.053438	0.217789	-0.04058	-0.11344	0.010211
Min	0		0.763	0.587	0.004	0.01	0.023	-0.055	-0.15	0.005
Max	0.416		0.993	0.988	0.032	0.091	0.826	-0.032	-0.09	0.017
Standard Deviation	0.122325		0.060393	0.105323	0.006642	0.021894	0.229922	0.00566	0.01568	0.003392
Confid. Interval	0.05361		0.026468	0.046159	0.002911	0.009595	0.100766	0.002481	0.006872	0.001487
Upper CI	0.175768		0.943468	0.901791	0.020595	0.063033	0.318555	-0.0381	-0.10657	0.011697
Lower CI	0.068548		0.890532	0.809473	0.014773	0.043842	0.117024	-0.04306	-0.12031	0.008724

AFB symptoms from Subsistence Farming SEM in Paper IV (cont.)

AFB on CornStorage			**AFB on PostHPrac**			**CornStorage on ROI**			**PostHPrac on Improved Qual**		
Unst. Est.	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**	**Unst. Est.**	**Stnd. Est.**	**P-value**
-0.017	-0.082	0.023	-0.027	-0.156	0.052	0.093	0.065	0.001	-1.242	-0.464	0
-0.015	-0.075	0.125	-0.021	-0.155	0.155	0.121	0.081	0.001	-1.242	-0.358	0
-0.012	-0.055	0.024	-0.034	-0.163	0.142	0.106	0.069	0.003	-1.171	-0.483	0
-0.015	-0.068	0.067	-0.026	-0.158	0.076	0.082	0.064	0.01	-1.319	-0.441	0.001
-0.018	-0.074	0.038	-0.026	-0.141	0.084	0.081	0.057	0.012	-1.208	-0.467	0
-0.019	-0.076	0.041	-0.03	-0.163	0.139	0.078	0.058	0.006	-1.084	-0.382	0.001
-0.023	-0.101	0.025	-0.027	-0.187	0.055	0.103	0.079	0.002	-1.445	-0.422	0
-0.013	-0.05	0.169	-0.043	-0.232	0.07	0.098	0.071	0.003	-1.173	-0.406	0
-0.019	0.079	0.031	-0.031	-0.129	0.065	0.094	0.06	0.003	-1.223	-0.558	0
-0.022	-0.091	0.033	-0.025	-0.157	0.118	0.095	0.063	0.006	-1.174	-0.35	0.001
-0.012	-	0.049	-0.023	-	0.057	0.087	0.064	0.007	-1.522	-0.444	0
-0.019	-0.068	0.043	-0.033	-0.18	0.051	0.084	0.068	0.005	-1.312	-0.416	0
-0.015	-	0.034	-0.034	-	0.059	0.08	0.053	0.011	-1.146	-0.505	0
-0.015	-	0.017	-0.028	-	0.1	0.11	0.082	0.001	-1.195	-0.414	0
-0.015	-0.075	0.065	-0.03	-0.166	0.111	0.085	0.056	0.011	-1.215	-0.47	0.001
-0.017	-0.069	0.037	-0.03	-0.17	0.107	0.074	0.055	0.021	-1.225	-0.371	0.001
-0.019	-0.074	0.089	-0.048	-0.209	0.1	0.111	0.073	0.001	-0.961	-0.338	0.013
-0.016	-0.054	0.152	-0.06	-0.243	0.09	0.088	0.065	0.006	-0.923	-0.345	0.011
-0.016	-0.083	0.026	-0.03	-0.176	0.084	0.076	0.061	0.007	-1.155	-0.563	0
-0.014	-0.063	0.085	-0.032	-0.147	0.08	0.098	0.069	0.003	-1.304	-0.564	0
-0.01653	-0.06231	0.060526	-0.03216	-0.1735	0.091737	0.092158	0.065684	0.006263	-1.21037	-0.43668	0.001526
-0.023	-0.101	0.017	-0.06	-0.243	0.051	0.074	0.053	0.001	-1.522	-0.564	0
-0.012	0.079	0.169	-0.021	-0.129	0.155	0.121	0.082	0.021	-0.923	-0.338	0.013
0.00308	0.039903	0.04451	0.009281	0.031226	0.03089	0.01338	0.008686	0.005031	0.140814	0.073428	0.003732
0.00135	0.017488	0.019507	0.004068	0.013685	0.013538	0.005864	0.003807	0.002205	0.061713	0.032181	0.001636
-0.01518	-0.04482	0.080033	-0.02809	-0.15981	0.105275	0.098022	0.069491	0.008468	-1.14866	-0.4045	0.003162
-0.01788	-0.0798	0.041019	-0.03623	-0.18719	0.078199	0.086294	0.061877	0.004058	-1.27208	-0.46886	-0.00011

Overall, results suggested that model fit indices, parameter estimates, and p-values within the SEMs were stable.

APPENDIX D. CLUSTERING ANALYSIS

In partnership with the Computer Engineering department a clustering algorithm exercise was also conducted on the USAID 2012 dataset. The algorithm attempted to cluster groups of children using variables as dividers. A description and preliminary results are presented below.

Introduction

In this report, we discuss analysis of the “USAID Data 2012 Final 1 Transformed – no specific nutrition.sav” data set. The goal of this exploratory data analysis is to identify distinct subgroups of individuals via cluster analysis techniques and identify important features to the selected clustering criteria. The initial data set consisted of 514 variables (features) and 5556 samples. The domain expert (Lee) assisted with reducing the number of features to 88 potentially important variables that were considered for the analysis. In this report, we describe the steps taken for data processing, clustering criteria and evaluation, and provide a set of 5 different clustering options for further investigation.

Data Processing

Data pre-processing consists of the following phases:

1. Elimination:

   To overcome the challenge of the missing values we eliminated any feature with more than 2% of missing values (i.e. has more than 100 missing). As a result, 3 variables (F12, B17 and HDDS) were removed in this process. Samples that have more than 3 missing values were also removed. After this process was complete, the updated data set contains 85 variables and 4715 samples.

2. Correlation Analysis:

   In this phase, we calculated the pairwise Pearson correlations between each of the features and filtered out features with a correlation greater than 0.75 with other variables. During this process, a total of 7 features were eliminated (Interv-date, zwaz, zwhz, sustainablelivestock, value-chain_cat, F10, water_treatment), leaving a total of 78 features.

3. Missing value replacement:

   The data set has 9 numerical features and the remaining 69 are categorical. We replaced the missing values in the numerical features by the average value of the feature before adding the missing value. That is, the missing values were replaced in order. After replacing one value, a new average was calculated which replaced the next missing value and so on. Thus, multiple missing values were not replaced with single value. For the categorical values, the missing values were replaced randomly with values from the same variable.

Clustering Criteria

The clustering algorithm we used is the k-dimensional sub-space clustering as described and discussed in the attached slides (power point slides attached). The algorithm consists of two phases: single-dimension and multi-dimension. In the single dimension clustering we classify the data samples based on each feature. In this case, we have 78 different clustering criteria. The features were ranked from best to worst based on the Silhouette index evaluation for each single dimension clustering. The top 60 features* were ranked as follows:

{ D54, diarrhea, ORTRecode, G16, H302_153, C16, H402_175, male_adult_max, G12, agemos, H501, E25, G49, G13_1CornDis, credit, F16, G14, I12Recode, improved_storage, E24, E20, zBMI, F08, zhaz, H302_151, H402_188, E16, G11_1CornFert, total_consumption, E21, I02Recode, G56, E12, E23, G18, G10, F09, F07, E22, poverty, H402_193, F11Sanitation, B18Educ, sex, E18, G39, H402_160, value_chain_any, E28, I17Recode, H402_190, production_plan, H302_142, G29, A06_1, E38, F04Recode, F15, E15}

These features have been divided into 4 levels and each level contains 15 feature. These features were used in the multi-dimensional clustering (also referred to as k-dimensional clustering), the algorithm groups the samples that have been assigned to the same clusters along each single dimension clustering to the same cluster. We obtained 123 different clustering criteria by using a moving window.

Cluster Evaluation

To evaluate the 123 different clustering criteria, we utilized two different approaches:

Internal Validation Indices:

- Davis-Bouldin (DB) Index: For each cluster C, the similarities between C and all other clusters are computed, and the highest value is assigned to C as its cluster similarity. Then the DB index can be obtained by averaging all the cluster similarities. The smaller the index, the better the clustering result. By minimizing this index, clusters are the most distinct from each other, and therefore achieves the best partition.

- Silhouette Index (SI): which validates the clustering performance based on the pairwise difference of between and within-cluster distances. The optimal cluster number is determined by maximizing the value of this index.

Statistical Evaluation:

- Nominal logistic regression was performed with the cluster number as the response and features as explanatory variables. Cluster membership can be predicted based on the model to determine how effective the features are in cluster classification. The model was built based on 2/3 of the data as a training set and 1/3 of the data was withheld as a test set. The classification error rate (CER) and r-square value in the test data set were calculated and used as additional criteria for cluster evaluation.

Clustering criteria that were in the lower quartile for Davis-Bouldin and upper quartile for Silhouette were chosen for further statistical evaluation with the logistic regression. The remaining clustering criteria were ranked based on their DB Index value, Silhouette Index value, test set R-square and CER. Values in the lower half for DB and CER were ranked “high” and values in the upper half for Silhouette and R-square were deemed “high”. Cluster criteria with the most “high” rankings across the four different criteria were ranked as the best.

Selected Clustering Criteria

Based on the validation indices and the statistical evaluation for clustering criteria, 5 were chosen for further investigation:

- Criteria 1, 2 and 3: The subspace selection features* are

{ **D54, diarrhea, ORTRecode, G16, H302_153, C16, H402_175, male_adult_max, G12, agemos, H501, E25, G49,G13_1CornDis, credit, F16, G14, I12Recode, improved_storage, E24, E20, zBMI, F08, zhaz, H302_151, H402_188, E16, G11_1CornFert, total_consumption**}.

In criteria 1 the allowed difference is 1, which means samples will be classified into the same cluster if they have been assigned to the same cluster along all the 15 single dimensional clusters except 1. Criteria 2 and 3 the allowed difference was 2 and 3 respectively.

- Criteria 4: All 78 features were involved in the subspace clustering but the allowed difference in 6.
- Criteria 5: the subspace selection features* are { **D54, diarrhea, ORTRecode, G16, H302_153, C16, H402_175, male_adult_max, G12, agemos, H501, E25, G49, G13_1CornDis, A06_1 F15 E15 E18 E28 E38 G29 G39 production_plan value_chain_any H302_142 H402_160 H402_190 F04Recode I17Recode**}. The allowed difference is 2.

*Note that one additional variable G05 was included, but it was the same value across all samples (no variation) and thus did not affect the clustering criteria.

Additional Files:

- Clustering Algorithm.ppt

  This file contains a description of the k-dimensional sub-space clustering algorithm.
- selected clustering criterias.xls

  This file contains the cluster labels for all samples for each of the selected clustering criteria.

Table 1. P-values for testing for significant differences among clusters.

Feature	**itr-31 (Cr1) CL4 rem.**	**Itr-32 (Cr2) CL3 rem.**	**Itr-33 (Cr3)**	**Itr115 (Cr4) CL3 rem.**	**Itr18 (Cr5) CL3 rem.**
D54	0.0167	0.1006	0.0061	0.6675	0.5504
diarrhea	0.0167	0.1006	0.0061	0.6675	0.5504
ORTRecode	0.0064	0.0372	0.0008	0.9	0.0025
G16	<0.0001	<0.0001	<0.0001	0.0003	<0.0001
H302_153	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
C16	<0.0001	<0.0001	<0.0001	0.6748	0.0003
H402_175,	<0.0001	<0.0001	<0.0001	0.0019	<0.0001
male_adult_max	0.5931	0.2669	0.4004	0.1202	0.0012
G12	<0.0001	<0.0001	<0.0001	0.0049	<0.0001
agemos	<0.0001	<0.0001	<0.0001	0.0006	<0.0001
H501	<0.0001	<0.0001	<0.0001	0.2746	<0.0001
E25	<0.0001	<0.0001	<0.0001	0.0128	<0.0001
G49	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
G13_1CornDis	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
credit	0.0018	0.0024	0.0006	0.0016	
F16	<0.0001	<0.0001	<0.0001	<0.0001	
G14	0.0059	0.0016	0.2483	<0.0001	
I12Recode	0.0015	0.0011	0.0005	<0.0001	
improved_storage	0.0017	0.0487	0.0005	0.0364	
E24	<0.0001	<0.0001	<0.0001	<0.0001	
E20	<0.0001	0.0001	<0.0001	0.0310	
zBMI	0.7267	0.9279	0.7600	0.3488	
F08	0.0029	0.0759	0.0030	0.0037	
F45	0.564	0.654	0.239	0.113	

Table 1. P-values for testing for significant differences among clusters (cont.)					
zhaz	0.7350	0.6883	0.7648	0.6979	
H302_151	0.0589	0.6097	0.4346	0.472	
H402_188	<0.0001	<0.0001	<0.0001	0.0029	
E16	0.0002	0.0031	0.0014	<0.0001	
G11_1CornFert	0.7351	0.4798	0.3990	<0.0001	
total_consumpt.	<0.0001	<0.0001	<0.0001	<0.0001	
A06_1				<0.0001	<0.0001
F15				<0.0001	0.0004
E15				<0.0001	0.0051
E18				0.0004	0.0221
E28				0.4514	0.7806
E38				0.484	0.0131
G29				0.6641	<0.0001
G39				0.778	<0.0001
production_plan				0.5706	0.0005
value_chain_any				0.2467	<0.0001
H302_142				<0.0001	0.0877
H402_160				0.0007	0.0255
H402_190				0.01	<0.0001
F04Recode				<0.0001	0.0007
I17Recode				<0.0001	0.9802

*For categorical variables, a chi-square test of association between cluster number and each variable is performed. P-values <0.05 indicate there is a significant association between cluster number and the variable.

*For quantitative variables (agemos, zBMI, zhaz, total_consumption, E38, wom_age, total_members, E37, Municipality), a one-way ANOVA is performed to test for a difference in means among clusters. P-values <0.05 indicate that there is a significant difference in the variable means somewhere amongst the clusters.

*The “outlier” cluster was removed for this testing since such extreme unbalance between sample numbers in the different clusters made it much more likely that the assumptions of the testing procedures would be violated.

*P-values that are highlighted in red correspond to variables that are both significant and exhibited the largest differences between clusters. For categorical variables, these represent a difference of at least 12% between clusters. Quantitative variables should be checked individually to see if the differences are practically meaningful.

*A PDF file of the JMP output is provided that provides graphs showing the distribution of the features among the clusters. For categorical variables, mosaic plots display percentages of observations that fall into each category for each cluster. For quantitative variables, plots of the variable verses cluster membership are given. You can check these files for more detailed information and to make sure I didn’t miss highlighting any significant variables that had large differences.

*Note that a multiple testing correction should be performed to reduce the probability of false positives among the set of tests. I did not do this for the data exploration purposes, but this can be done for publication purposes.

*Note that the women’s age variable analysis for Itr115 had some outliers with ages over 900. I removed these and rechecked the test, which is why you see two pvalues in the table for that variable. Pvalue with the outliers was 0.008 and without the outliers was <0.001. The output without the outliers is not given in the PDF file, but I can send it separately if needed.

Selected Clustering Criteria with Significant zhaz and diarrhea

In order to further examine different clustering criteria, additional testing was conducted to test for significant differences between clusters on the zhaz and diarrhea variables. The DB and SI validation indices were utilized to select 2 clustering criteria from among all of those with significant differences in zhaz and/or diarrhea between clusters. (Table 3 provides further information):

- The subspace selection features* for both of the additional criteria are:

{ **D54, diarrhea, ORTRecode, G16, H302_153, C16, H402_175, male_adult_max, G12, agemos, H501, E25, G49,G13_1CornDis, credit, F16, G14, I12Recode, improved_storage, E24, E20, zBMI, F08, zhaz, H302_151, H402_188, E16, G11_1CornFert, total_consumption. E21, I02Recode, G56, E12, E23, G18, G10, F09, F07, E22, poverty, H402_193, F11Sanitation, B18Educ, sex**}.

In criteria 1 the allowed difference is 2; whereas the allowed difference is 4 in criteria 2.

*Note that one additional variable G05 was included, but it was the same value across all samples (no variation) and thus did not affect the clustering criteria.

Table 2. Selected Clustering Criteria with significant zhaz and/or diarrhea

Criteria name	**itr-60 (Cr1)**	**itr-62(Cr2)**
Davis_Bouldin	532.7135	294.9817
Silhouette	0.053048	0.054975
# of clusters	4	3
# samples in cluster1	4260	4573
# samples in cluster2	390	140
# samples in cluster3	63	1
# samples in cluster4	1	0
# of features in subspace	46	46
Allowed difference	2	4

Table 3. P-values for testing for significant differences among clusters.

Feature	**itr-60 (Cr1) CL4 rem.**	**Itr-62 (Cr2) CL3 rem.**
D54	0.0062	0.0071
diarrhea	0.0062	0.0071
ORTRecode	0.3701	0.2366
G16	<0.0001	<0.0001
H302_153	<0.0001	0.0001

Table 3. P-values for testing for significant differences among clusters (cont.)

C16	0.1605	0.3730
H402_175	<0.0001	0.0017
male_adult_max	0.3165	0.1522
G12	<0.0001	<0.0001
agemos	<0.0001	<0.0001
H501	0.0011	0.0036
E25	<0.0001	<0.0001
G49	<0.0001	<0.0001
G13_1CornDis	<0.0001	<0.0001
credit	0.2386	0.8252
F16	<0.0001	<0.0001
G14	<0.0001	<0.0001
I12Recode	<0.0001	<0.0001
improved_storage	0.5309	0.2517
E24	<0.0001	0.0006
E20	0.0537	0.0346
zBMI	<0.0001	<0.0001
F08	<0.0001	0.0380
zhaz	0.1377	0.0179
H302_151	<0.0001	<0.0001
H402_188	<0.0001	<0.0001
E16	<0.0001	<0.0001
G11_1CornFert	<0.0001	<0.0001
total_consumpt.	<0.0001	<0.0001
E21	<0.0001	<0.0001
I02Recode	<0.0001	<0.0001
G56	<0.0001	<0.0001
E12	<0.0001	0.0004
E23	<0.0001	<0.0001

Table 3. P-values for testing for significant differences among clusters (cont.)		
G18	<0.0001	<0.0001
G10	<0.0001	<0.0001
F09	<0.0001	0.0072
F07	0.0009	0.0017
E22	<0.0001	0.0008
poverty	<0.0001	<0.0001
H402_193	<0.0001	<0.0001
F11Sanitation	<0.0001	<0.0001
B18Educ	<0.0001	<0.0001
sex	<0.0001	<0.0001

*The same notes from the previous analyses (Table 1) apply here as well.

*P-values that are highlighted in red correspond to variables that are both significant and represent a difference of at least 12% between the main two clusters (for categorical variables). Quantitative variables should be checked individually to see if the differences are practically meaningful.

From Table 1 above, below is a list of the top 13 variables which were most commonly significant across the 5 different sets of selected variables for improved child grouping. No specific focus on HAZ/diarrhea was given. The primary theme of the variables included *socio economic inputs* (e.g. maize cultivation practices) and <u>socio economic outputs</u> (e.g. spent $$ on medications).

1. *Maize cultivation practices*
2. Food – sweets and chocolates
3. *Soil conservation practices*
4. *Problem with diseases in the maize*
5. <u>Spent $$ on medications in past month</u>
6. *Household saved maize harvest*
7. <u>Spent $$ on medical tests</u>
8. Age of the child
9. <u>Household owns the house lived in</u>

10. Food – candies
11. <u>Total consumption scored based on owned items</u>
12. Presence of soap at handwashing station
13. <u>Spent $$ on school enrollment</u>

From Table 2 above, below is a list of the top 15 variables which had the largest contribution to creating groups/clusters of children where HAZ and diarrhea were also used. Top 10 potentially related with *economics*, <u>nutrition</u>, **education**, and hygiene.

1. *Poverty index (ownership of items)*
2. <u>Food – oil, butter, margarine</u>
3. *Household Practices Soil Conservation*
4. *Household bred animals last year*
5. **Mother knows warning signs for problems in pregnancy**
6. Presence of soap at handwashing station
7. **Mother knows warning signs for problems with sick child**
8. <u>Food – other fruits and veggies</u>
9. <u>Food – sweets and chocolates</u>
10. *Household spent money on electricity*
11. Household used potentially harmful fertilizers
12. Household was devoted to the cultivation of beans
13. Household has problems with disease, pests or weather in maize cultivation
14. Did the mother attend school
15. Household has problems with maize cultivation

Options for next steps to have a strong enough publication for Science or Nature;

1) Conduct same analysis on additional dataset, publication focused on stunting outcome
2) Compare with traditional clustering algorithm, publication focused on methodology
3) Conduct SEM on top 10 variables, publication focused more on methods then stunting

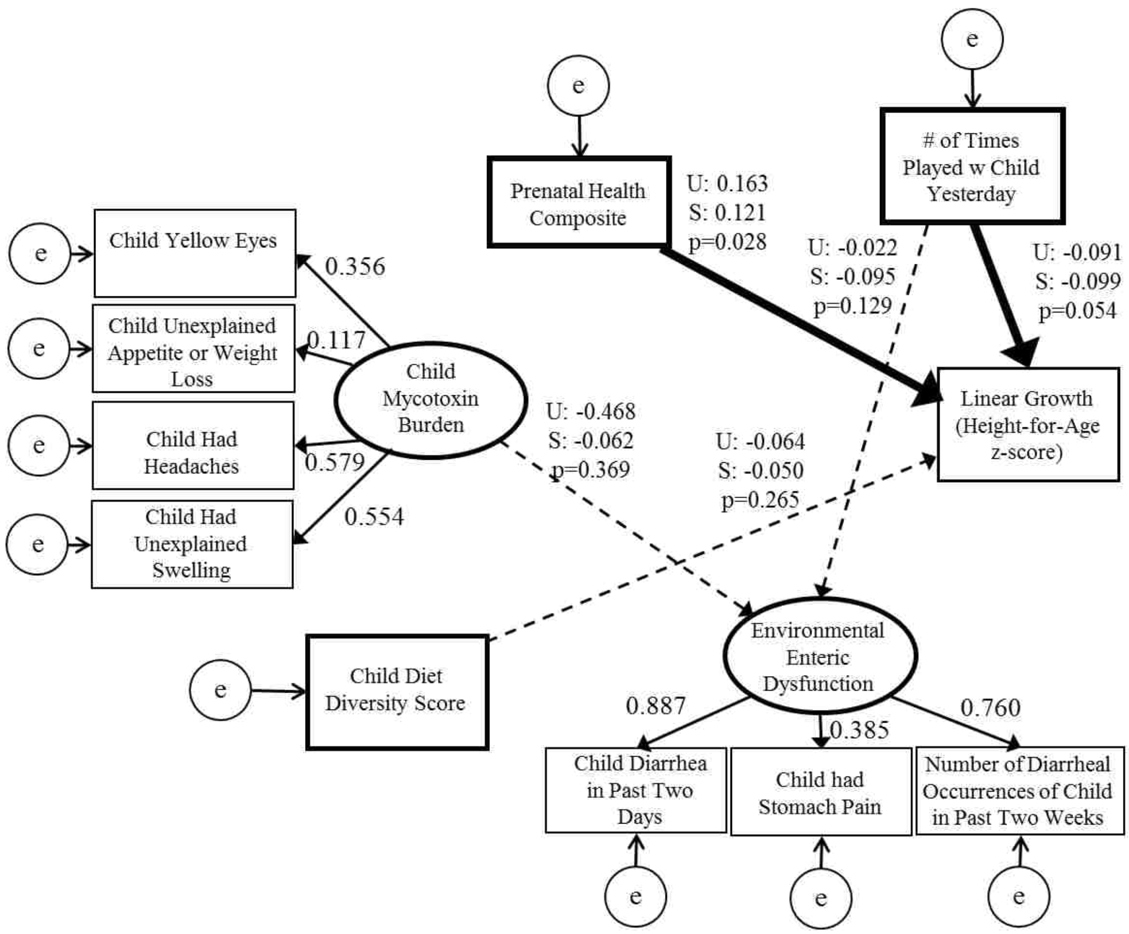
APPENDIX E. ADDITIONAL SEMS

Additional SEMs were hypothesized and tested, but were not part of the main hypotheses or objectives of the dissertation, but do provide insight. These are presented below.

Child height-for-age SEMs

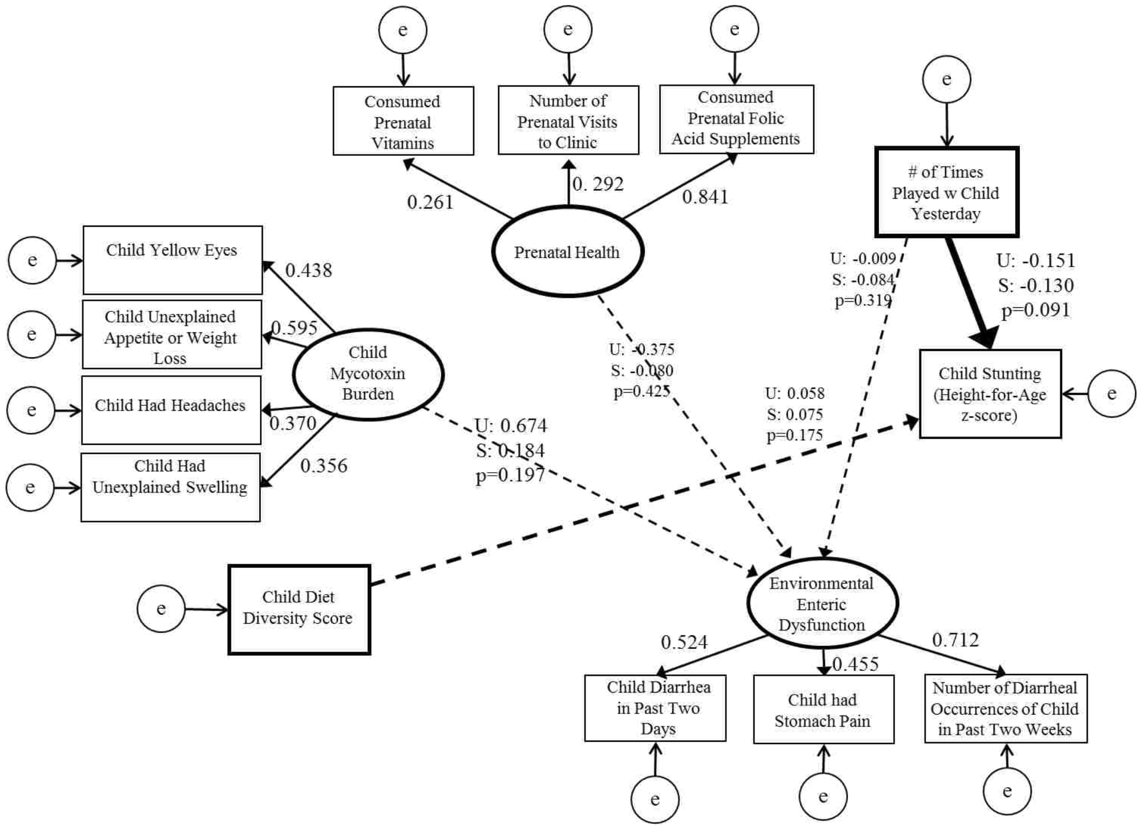
October 2016 data using a composite variable for prenatal health.

Used DWLS robust estimator; N=372; Chi-square: 52.988, p=0.034; RMSEA: 0.033 (CI: 0.010 – 0.051); Robust CFI: 0.904; Robust TLI: 0.861



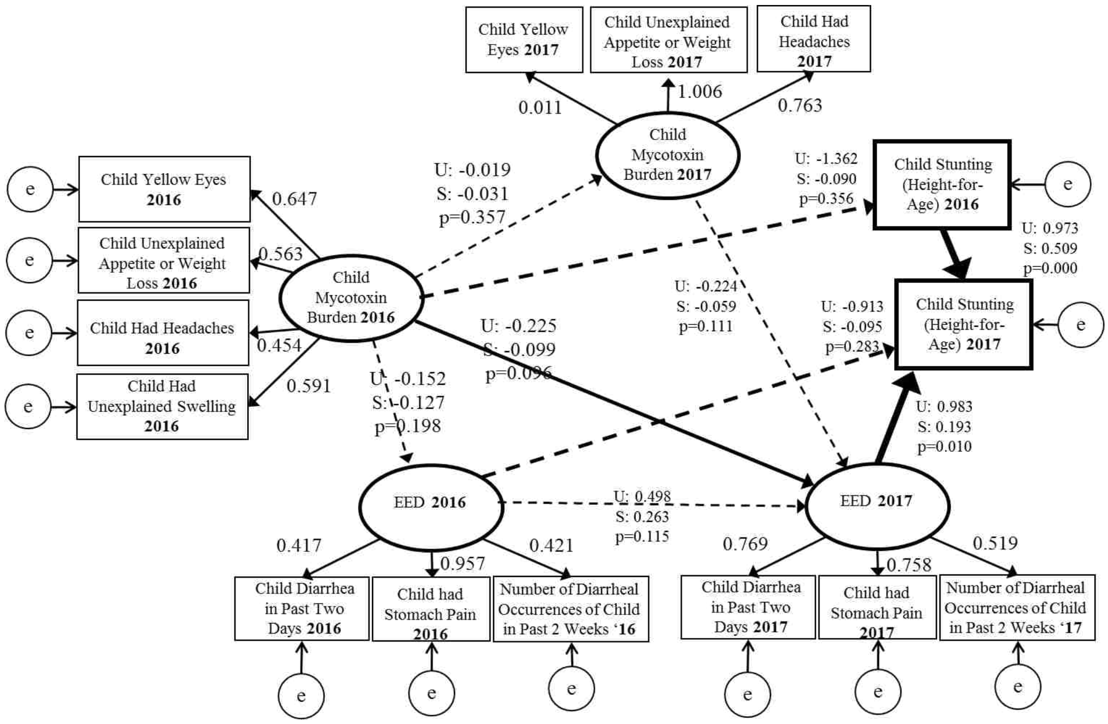
February 2017 full SEM with a latent variable for prenatal health

Used DWLS robust estimator; N=300; Chi-square: 73.848, p=0.078; RMSEA: 0.026 (CI: 0.000 – 0.042); Robust CFI: 0.855; Robust TLI: 0.808



EED, AFB, and HAZ over time SEM

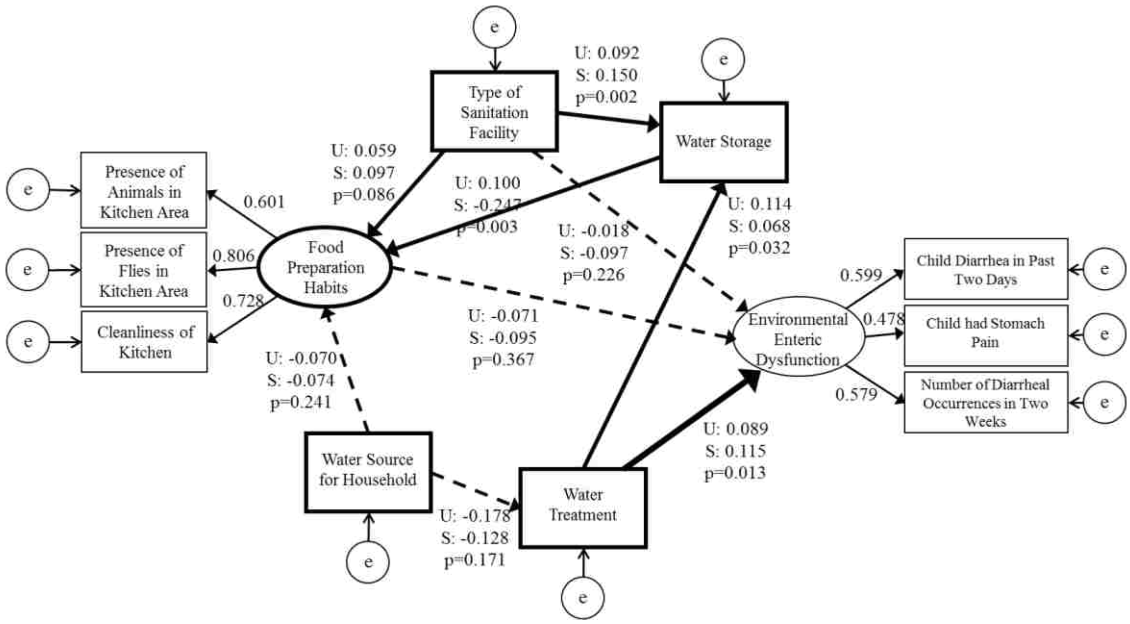
Used DWLS robust estimator; N=146; Chi-square: 84.134, p=0.270; RMSEA: 0.021 (CI: 0.000 – 0.045); Robust CFI: 0.995; Robust TLI: 0.939



EED SEMs

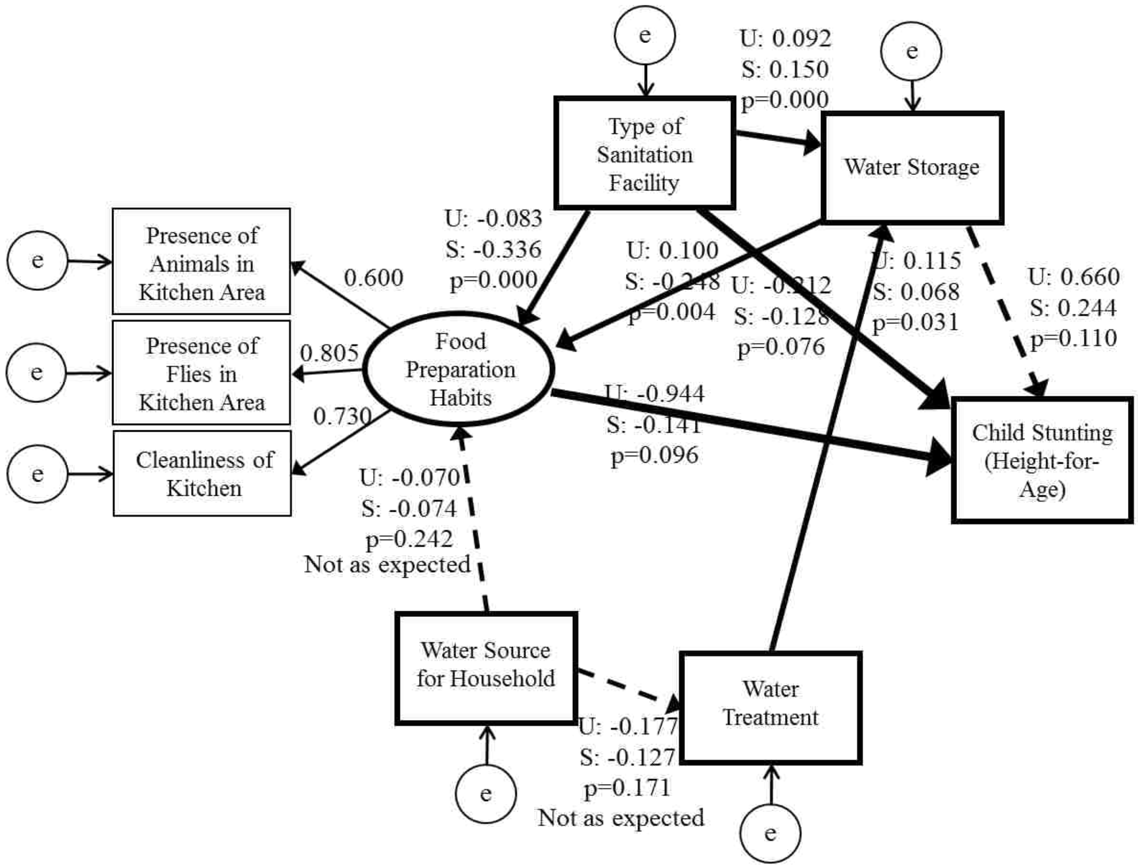
February 2017 data applied to the SEM

Used DWML robust estimator; N=310; Chi-square: 26.564, p=0.432; RMSEA: 0.007 (CI: 0.000 – 0.037); Robust CFI: 0.999; Robust TLI: 0.998



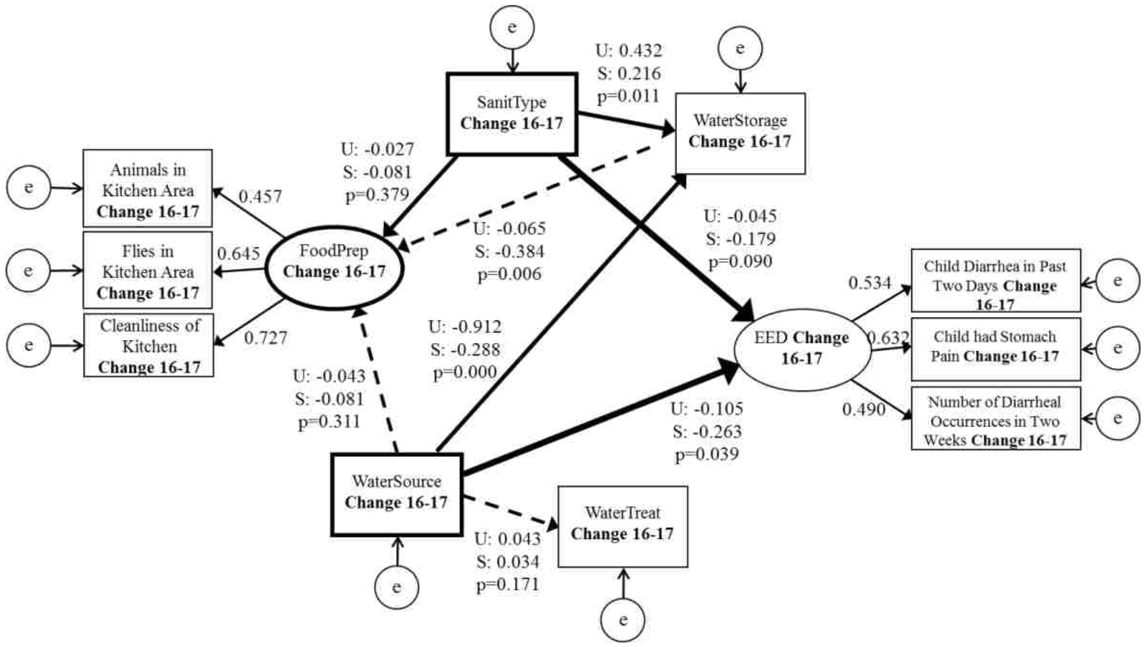
February 2017 WASH model with HAZ as an outcome

Used DWML robust estimator; N=310; Chi-square: 10.631, p=0.642; RMSEA: 0.000 (CI: 0.000 – 0.036); Robust CFI: 1.000; Robust TLI: 1.008



Change over time among all variables

Used DWLS robust estimator; N=153; Chi-square: 34.877, p=0.090; RMSEA: 0.043 (CI: 0.000 – 0.074); Robust CFI: 0.953; Robust TLI: 0.917



REFERENCES

(1) United Nations General Assembly. *Transforming our world: the 2030 Agenda for Sustainable Development*; 2015; Vol. A/70/L.1, pp 1–35.

(2) World Health Organization. *Country Profile Indicators: Interpretation Guide*; Geneva, Switzerland, 2010.

(3) Alderman, H.; Hodditnott, J.; Kinsey, B. Long-term Consequences of Early Childhood Malnutrition. *Oxf. Econ. Pap.* **2006**, *58* (168), 450–474.

(4) Dewey, K. G.; Begum, K. Long-term consequences of stunting in early life. *Matern. Child Nutr.* **2011**, *7* (SUPPL. 3), 5–18.

(5) United Nations Children's Emergency Fund. *Improving Child Nutrition: The achievable imperative for global progress*; 2013.

(6) UNICEF. Malnutrition: Current Status + Progress http://data.unicef.org/nutrition/malnutrition.html.

(7) Onyango, A. W. Promoting healthy growth and preventing childhood stunting: A global challenge. *Matern. Child Nutr.* **2013**, *9* (S2), 1–5.

(8) Esrey, S. A. Water , Waste , and Well-Being : A Multicountry Study. *Am. J. Epidemiol.* **1996**, *143* (6), 608–623.

(9) Wild, C. P.; Miller, J. D.; Groopman, J. D. *Mycotoxin Control in Low- and Middle- Income Countries*; 2016.

(10) Solomons, N. W.; Vossenaar, M.; Chomat, A.; Doak, C. M.; Koski, K. G.; Scott, M. E. Stunting at birth: recognition of early-life linear growth failure in the western highlands of Guatemala. *Public Health Nutr.* **2014**, *18* (10), 1737–1745.

(11) Berngard, S. C.; Berngard, J. B.; Krebs, N. F.; Garcés, A.; Miller, L. V.; Westcott, J.; Wright, L. L.; Kindem, M.; Hambidge, K. M. Newborn length predicts early infant linear growth retardation and disproportionately high weight gain in a low-income population. *Early Hum. Dev.* **2013**, *89* (12), 967–972.

(12) Mwaniki, E. W.; Makokha, a N. Nutrition status and associated factors among children in public primary schools in Dagoretti, Nairobi, Kenya. *Afr. Health Sci.* **2013**, *13* (1), 39–46.

(13) Khor, G. L.; Sharif, Z. M. Dual forms of malnutrition in the same households in Malaysia - A case study among Malay rural households. *Asia Pac. J. Clin. Nutr.* **2003**, *12* (4), 427–437.

(14) 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.; et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet* **2013**, *382* (9890), 427–451.

(15) MSPAS/INE/CDC. *Encuesta Nacional de Salud Materno Infantil 2008 (ENSMI) 2008/2009*; Guatemala, 2010.

(16) Keusch, G. T.; Rosenberg, I. H.; Denno, D. M.; Duggan, C.; Guerrant, R. L.; Lavery, J. V.; Tarr, P. I.; Ward, H. D.; Black, R. E.; Nataro, J. P.; et al. Implications of acquired environmental enteric dysfunction for growth and stunting in infants and children living in low-and middle-income countries. *Food Nutr. Bull.* **2013**, *34* (3), 357–364.

(17) Korpe, P. S.; Petri, W. a. Environmental enteropathy: Critical implications of a poorly understood condition. *Trends Mol. Med.* **2012**, *18* (6), 328–336.

(18) International Agency for Research on Cancer. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Aflatoxins*; 2006.

(19) Checkley, W.; Buckley, G.; Gilman, R. H.; Assis, A. M.; Guerrant, R. L.; Morris, S. S.; Mølbak, K.; Valentiner-Branth, P.; Lanata, C. F.; Black, R. E. Multi-country analysis of the effects of diarrhoea on childhood stunting. *Int. J. Epidemiol.* **2008**, *37* (4), 816–830.

(20) Ahmed, T.; Auble, D.; Berkley, J. a; Black, R.; Ahern, P. P.; Hossain, M.; Hsieh, A.; Ireen, S.; Arabi, M.; Gordon, J. I. An evolving perspective about the origins of childhood undernutrition and nutritional interventions that includes the gut microbiome. *Ann. N. Y. Acad. Sci.* **2014**, *1332*, 22–38.

(21) Humphrey, J. H. Child undernutrition, tropical enteropathy, toilets, and handwashing. *Lancet* **2009**, *374* (9694), 1032–1035.

(22) Klaver, W. *Underweight or stunting as an indicator of the MDG on poverty and hunger*; Leiden, The Netherlands, 2010.

(23) Guerrant, R. L.; Oriá, R. B.; Moore, S. R.; Oriá, M. O. B.; Lima, A. A. A. M. Malnutrition as an enteric infectious disease with long-term effects on child development. *Nutr. Rev.* **2008**, *66* (9), 487–505.

(24) Kau, A. L.; Ahern, P. P.; Griffin, N. W.; Goodman, A. L.; Gordan, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **2011**, *474*, 327–336.

(25) Martorell, R.; Kettel Khan, L.; Schroeder, D. G. Reversibility of stunting: epidemiological findings in children from developing countries. *Eur. J. Clin. Nutr.* **1994**, *48 (Suppl.*, S45–S57.

(26) Blanton, L. V; Charbonneau, M. R.; Salih, T.; Barratt, M. J.; Venkatesh, S.; Ilkaveya, O.; Subramanian, S.; Manary, M. J.; Trehan, I.; Jorgensen, J. M.; et al. Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science (80-.).* **2016**, *351* (6275).

(27) Schwarzer, M.; Makki, K.; Storelli, G.; Machuca-gayet, I.; Hudcovic, T.; Heddi, A.; Rieusset, J.; Kozakova, H. Lactobacillus plantarum strain maintains growth of infant mice during chronic undernutrition. *Science (80-.).* **2016**, *351* (6275), 854–857.

(28) USAID. *Formative Research: Key Influencers of Household Food Access in the Western Highlands of Guatemala*; Washington, DC, 2015.

(29) Gibson, R. S. Zinc: the missing link in combating micronutrient malnutrition in developing countries. *Proc. Nutr. Soc.* **2006**, *65*, 51–60.

(30) Ziegler, E. E. Nutrient Needs for Catch-Up Growth in Low-Birthweight Infants. *Nestle Nutr. Work.* **2015**, *81*, 135–143.

(31) Viswanathan, V. K.; Hodges, K.; Hecht, G. Enteric infection meets intestinal function: how bacterial pathogens cause diarrhoea. *Nat. Rev. Microbiol.* **2009**, *7* (2), 110–119.

(32) WHO. *MCEE-WHO methods and data sources for child causes of death 2000-2015*; Geneva, Switzerland, 2016.

(33) Gerson, C. D.; Kent, T. H.; Lindenbaum, J. Recovery of small-intestinal structure and function after residence in the tropics. I. Studies in Peace Corps Volunteers. *Ann. Intern. Med.* **1971**, *74*, 218–222.

(34) Center for Disease Control and Prevention. *Diarrhea: Common Illness, Global Killer*; 2013.

(35) The World Bank. Health Benefits - Sanitation and Hygiene: Why They Matter http://water.worldbank.org/shw-resource-guide/sanitation-and-hygiene-why-they-matter/health-benefits.

(36) Moll, D. M.; McElroy, R. H.; Sabogal, R.; Corrales, L. F.; Gelting, R. J. Health impact of water and sanitation infrastructure reconstruction programmes in eight Central American communities affected by Hurricane Mitch. *J. Water Health* **2007**, *5* (1), 51–65.

(37) Vasquez, W. F.; Aksan, A. M. Water, sanitation, and diarrhea incidence among children: Evidence from Guatemala. *Water Policy* **2015**, *17* (5), 932–945.

(38) Reller, M. E.; Mendoza, C. E.; Lopez, M. B.; Alvarez, M.; Hoekstra, R. M.; Olson, C. A.; Baier, K. G.; Keswick, B. H.; Luby, S. P. A randomized controlled trial of household-based flocculant-disinfectant drinking water treatment for diarrhea prevention in rural Guatemala. *Am. J. Trop. Med. Hyg.* **2003**, *69* (4), 411–419.

(39) Rosa, G.; Miller, L.; Clasen, T. Microbiological effectiveness of disinfecting water by boiling in rural Guatemala. *Am. J. Trop. Med. Hyg.* **2010**, *82* (3), 473–477.

(40) Kaminsky, J. A.; Javernick-Will, A. N. The internal social sustainability of sanitation infrastructure. *Environ. Sci. Technol.* **2014**, *48* (17), 10028–10035.

(41) Goldman, N.; Pebley, A. R.; Beckett, M. Diffusion of ideas about personal hygiene and contamination in poor countries: Evidence from Guatemala. *Soc. Sci. Med.* **2001**, *52* (1), 53–69.

(42) Webb, A. L.; Stein, A. D.; Ramakrishnan, U.; Hertzberg, V. S.; Urizar, M.; Martorell, R. A simple index to measure hygiene behaviours. *Int. J. Epidemiol.* **2006**, *35* (6), 1469–1477.

(43) Center for Disease Control and Prevention. Water, Sanitation, & Hygiene (WASH)-related Emergencies & Outbreaks http://www.cdc.gov/healthywater/emergency/extreme-weather/floods-standingwater.html.

(44) Engebretsen, I. M. S.; Jackson, D.; Fadnes, L. T.; Nankabirwa, V.; Diallo, A. H.; Doherty, T.; Lombard, C.; Swanvelder, S.; Nankunda, J.; Ramokolo, V.; et al. Growth effects of exclusive breastfeeding promotion by peer counsellors in sub-Saharan Africa: the cluster-randomised PROMISE EBF trial. *BMC Public Health* **2014**, *14*, 633.

(45) Beusekom, I. Van; Vossenaar, M.; Montenegro-bethancourt, G.; Colleen, M.; Solomons, N. W. Estimates of exclusive breastfeeding rates among mother – infant dyads in Quetzaltenango , Guatemala , vary according to interview method and time frame. *United Nations Univ.* **2013**, *34* (2), 160–168.

(46) WHO; UNICEF. WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation http://www.wssinfo.org/.

(47) Torres, O.; Matute, J.; Gelineau-van Waes, J.; Maddox, J. R.; Gregory, S. G.; Ashley-Koch, a. E.; Showker, J. L.; Voss, K. a.; Riley, R. T. Human health implications from co-exposure to aflatoxins and fumonisins in maize-based food in Latin America: Guatmeala as a Case Study. *World Mycotoxin J.* **2015**, *8* (2), 143–159.

(48) Williams, J. H.; Phillips, T. D.; Jolly, P. E.; Stiles, J. K.; Jolly, C. M.; Aggarwal, D. Human aflatoxicosis in developing countries: a review of toxicology, exposure, potential health consequences, and interventions. *Am. Soc. Clin. Nutr.* **2004**, *80* (June 2015), 1106–1122.

(49) U.S. Food and Drug Administration. *CPG Sec. 555.400 Foods - Adulteration with Aflatoxin*; United States, 1980.

(50) Marasas, W. F. O.; Riley, R. T.; Hendricks, K. A.; Stevens, V. L.; Sadler, T. W.; Waes, J. G.; Missmer, S. A.; Cabrera, J.; Torres, O.; Gelderblom, W. C. A.; et al. Fumonisins Disrupt Sphingolipid Metabolism , Folate Transport , and Neural Tube Development in Embryo Culture and In Vivo : A Potential Risk Factor for Human Neural Tube Defects among Populations Consuming. *J. Nutr.* **2004**, *134*, 711–716.

(51) Kimanya, M. E.; De Meulenaer, B.; Roberfroid, D.; Lachat, C.; Kolsteren, P. Fumonisin exposure through maize in complementary foods is inversely associated with linear growth of infants in Tanzania. *Mol. Nutr. Food Res.* **2010**, *54* (11), 1659–1667.

(52) Smith, L. E.; Stoltzfus, R. J.; Prendergast, a. Food Chain Mycotoxin Exposure, Gut Health, and Impaired Growth: A Conceptual Framework. *Adv. Nutr. An Int. Rev. J.* **2012**, *3* (4), 526–531.

(53) Castelino, J. M.; Routledge, M. N.; Wilson, S.; Dunne, D. W.; Mwatha, J. K.; Gachuhi, K.; Wild, C. P.; Gong, Y. Y. Aflatoxin exposure is inversely associated with IGF1 and IGFBP3 levels in vitro and in Kenyan school children. *Mol. Nutr. Food Res.* **2015**, *59* (3), 574–581.

(54) United States Geological Survey. *Production and Market Flow Map: Guatemala First Season Maize*; 2013.

(55) United States Geological Survey. *Production and Market Flow Map: Guatemala Second Season Maize*; 2013.

(56) Mercy Corps Guatemala. *Barrier Analysis on Maize Harvest and Post-Harvest Practices*; 2015.

(57) Wu, F.; Bhatnagar, D.; Bui-Klimke, T.; Carbone, I.; Hellmich, R.; Munkvold, G.; Paul, P.; Payne, G.; Takle, E. Climate change impacts on mycotoxin risks in US maize. *World Mycotoxin J.* **2011**, *4* (1), 79–93.

(58) Rah, J. H.; Christian, P.; Shamim, A. A.; Arju, U. T.; Labrique, A. B.; Rashid, M. Pregnancy and lactation hinder growth and nutritional status of adolescent girls in rural Bangladesh. *J. Nutr.* **2008**, *138* (8), 1505–1511.

(59) Uauy, R.; Kain, J.; Mericq, V.; Rojas, J.; Corvalan, C. Nutrition, child growth, and chronic disease prevention. *Ann. Med.* **2008**, *40* (1), 11–20.

(60) Boeing, H. Nutritional epidemiology: New perspectives for understanding the diet-disease relationship? *Eur. J. Clin. Nutr.* **2013**, *67* (5), 424–429.

(61) Guerrant, R. L.; Schorling, J. B.; McAuliffe, J. F.; De Souza, M. a. Diarrhea as a cause and an effect of malnutrition: Diarrhea prevents catch-up growth and malnutrition increases diarrhea frequency and duration. *American Journal of Tropical Medicine and Hygiene*. 1992, pp 28–35.

(62) Villar, J.; Papageorghiou, A. T.; Pang, R.; Ohuma, E. O.; Ismail, L. C.; Barros, F. C.; Lambert, A.; Carvalho, M.; Jaffer, Y. a; Bertino, E.; et al. The likeness of fetal growth and newborn size across non-isolated populations in the INTERGROWTH-21(st) Project: the Fetal Growth Longitudinal Study and Newborn Cross-Sectional Study. *lancet. Diabetes Endocrinol.* **2014**, *8587* (14), 1–12.

(63) Neufeld, L. M.; Haas, J. D.; Grajéda, R.; Martorell, R. Changes in maternal weight from the first to second trimester of pregnancy are associated with fetal growth and infant length at birth. *Am. J. Clin. Nutr.* **2004**, *79* (4), 646–652.

(64) Kramer, M. S. Determinants of low birth weight: methodological assessment and meta-analysis. *Bull. World Health Organ.* **1987**, *65* (5), 663–737.

(65) Turner, P. C.; Collinson, A. C.; Cheung, Y. B.; Gong, Y.; Hall, A. J.; Prentice, A. M.; Wild, C. P. Aflatoxin exposure in utero causes growth faltering in Gambian infants. *Int. J. Epidemiol.* **2007**, *36* (5), 1119–1125.

(66) World Health Organization. *Counselling for Maternal and Newborn Health Care*; Geneva, Switzerland, 2013.

(67) Brown, J. E.; Murtaugh, M. A.; Jacobs, D. R.; Margellos, H. C. Variation in newborn size according to pregnancy weight change by trimester. *Am. J. Clin. Nutr.* **2002**, *76* (1), 205–209.

(68) Sekiya, N.; Anai, T.; Matsubara, M.; Miyazaki, F. Maternal Weigth Gain Rate in the Second Trimester are Associated with Birth Weight and Length of Gestation. *Gynecol. Obstet. Invest.* **2007**, *63* (1), 45–48.

(69) Strauss, R. S.; Dietz, W. H. Low maternal weight gain in the second or third trimester increases the risk for intrauterine growth retardation. *J. Nutr.* **1999**, *129*, 988–993.

(70) World Health Organization. Media Centre: Maternal Mortality http://www.who.int/mediacentre/factsheets/fs348/en/.

(71) Martorell, R.; Zongrone, A. Intergenerational influences on child growth and undernutrition. *Paediatr. Perinat. Epidemiol.* **2012**, *26* (SUPPL. 1), 302–314.

(72) Ramakrishnan, U.; Martorell, R.; Schroeder, D. G.; Flores, R. Role of intergenerational effects on linear growth. *J. Nutr.* **1999**, *129* (2S Suppl), 544S – 549S.

(73) Nix, J. “Invisible” crisis scars children for life. *CNN*. 2015.

(74) Forero-Ramirez, N.; Gamboa, L. F.; Bedi, A.; Sparrow, R. Child malnutrition and prenatal care: evidence from three Latin American countries. *Pan Am. J. public Heal.* **2014**, *35* (3), 163–171.

(75) Prendergast, A. J.; Humphrey, J. H. The stunting syndrome in developing countries. **2014**, 1–16.

(76) Publica, M. de S. *Encuesta Nacional de Salud Materno Infantil 2008-2009*; 2010.

(77) Ulijaszek, S. J. Relationships between undernutrition, infection, and growth and development. *Hum. Evol.* **1996**, *11* (3), 233–248.

(78) USAID; Food and Nutrition Technical Assistance III. *Food-Based Approaches to Improve the Quality and Diversity of Diets in the Western Highlands of Guatemala*; 2015.

(79) Food and Agriculture Organization. *An Introduction to the Basic Concepts of Food Security*; 2008.

(80) Rivera, J. a.; Hotz, C.; Gonzalez-Cossio, T.; Neufeld, L. M.; Garcia-Guerra, A. The Effect of Micronutrient Deficiences on Child Growth: A Review of Results from Community-Based Supplementation Trials. *J. Nutr.* **2003**, *133*, 4010S – 4020S.

(81) Georgieff, M. K. Nutrition and the developing brain: Nutrient priorities and measurement. *Am. J. Clin. Nutr.* **2007**, *85 (suppl)*, 614S – 620S.

(82) Miller, M. The MAL-ED study: A multinational and multidisciplinary approach to understand the relationship between enteric pathogens, malnutrition, gut physiology, physical growth, cognitive development, and immune responses in infants and children up to 2 years of . *Clin. Infect. Dis.* **2014**, *59* (Suppl 4), S193–S206.

(83) Shugart, B. J. MOTHER LODE Superhero sugars in breast milk make the newborn gut safe for beneficial bacteria. *Sci. News* **2016**, *185* (1), 22–26.

(84) WHO. Guiding Principles for Complementary Feeding of the Breastfed Child. **2013**.

(85) Bhandari, N.; Mazumder, S.; Bahl, R.; Martines, J.; Black, R. E.; Bhan, M. K. An Educational Intervention to Promote Appropriate Complementary Feeding Practices and Physical Growth in Infants and Young Children in Rural Haryana, India. *J. Nutr.* **2004**, *134*, 2342–2348.

(86) UNICEF. At a Glance: Guatemala Statistics http://www.unicef.org/infobycountry/guatemala_statistics.html.

(87) Chary, A.; Messmer, S.; Sorenson, E.; Henretty, N.; Dasgupta, S.; Rohloff, P. The Normalization of Childhood Disease: An Ethnographic Study of Child Malnutrition in Rural Guatemala. *Hum. Organ.* **2013**, *72* (2), 87–97.

(88) Steinman, C.; Deshpande, R.; Farley, J. . . Understanding Ethnic Variation in Pregnancy-related Care in Rural Guatemala. *Ethn. Health* **2000**, *5* (1), 5–22.

(89) UNICEF; Liverpool School of Tropical Medicine. *Gender Influences on Child Survival , Health and Nutrition : A Narrative Review*; 2011.

(90) Ozaltin, E.; Hill, K.; Subramanian, S. V. Association of maternal stature with offspring mortality, underweight, and stunting in low- to middle-income countries. *JAMA* **2010**, *303* (15), 1507–1516.

(91) Hernández-Díaz, S.; Peterson, K. E.; Dixit, S.; Hernández, B.; Parra, S.; Barquera, S.; Sepúlveda, J.; Rivera, J. a. Association of maternal short stature with stunting in Mexican children: common genes vs common environment. *Eur. J. Clin. Nutr.* **1999**, *53* (12), 938–945.

(92) Sage, A. P.; Armstrong Jr., J. E. *Introduction to Systems Engineering*; Wiley, 2000.

(93) Kline, R. B. *Principles and Practices of Structural Equation Modeling*, Second.; The Guilford Press: New York, 2005.

(94) Grace, J. B. *Structural Equation Modeling and Natural Systems*, First.; Cambridge University Press: United Kingdom, 2006.

(95) Wu, X.; Kumar, V.; Ross, Q. J.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G. J.; Ng, A.; Liu, B.; Yu, P. S.; et al. *Top 10 algorithms in data mining*; 2008; Vol. 14.

(96) Ford, A. *Modeling the Environment*, 2nd ed.; Island Press: Washington, DC, 2010.

(97) Albin, S.; Forrester, J. *Building a System Dynamics Model*; 1997.

(98) Hovmand, P. S. Analyzing dynamic systems: a comparison of structural equation modeling and system dynamics modeling. In *Structural Equation Modeling: Applications in Ecological and Evolutionary Biology*; 2003; pp 212–234.

(99) Divelbiss, D. W.; Boccelli, D. L.; Succop, P. A.; Oerther, D. B. Environmental health and household demographics impacting biosand filter maintenance and diarrhea in Guatemala: an application of structural equation modeling. *Environ. Sci. Technol.* **2013**, *47* (3), 1638–1645.

(100) Voth-gaeddert, L. E.; Divelbiss, D. W.; Oerther, D. B. Utilizing Structural Equation Modeling to Correlate Biosand Filter Performance and Occurrence of Diarrhea in the Village of Enseado do Aritapera in Para, Brazil. *Water Sci. Technol. Water Supply* **2015**, *15* (1), 164–172.

(101) Voth-Gaeddert, L. E.; Divelbiss, D. W.; Oerther, D. B. Utilizing Structural Equation Modeling as an Evaluation Tool for Critical Parameters of the Biosand Filter in a Pilot Study in Para, Brazil. *Water Sci. Technol. Water Supply* **2015**, *15* (4), 845–851.

(102) Voth-Gaeddert, L. E.; Oerther, D. B. Measuring multidimensional poverty in a complex environment ; identifying the sensitive links. *Procedia Eng.* **2015**, *107*, 172–180.

(103) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer, 2009.

(104) James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, 2013.

(105) Aday, L. A.; Cornelius, L. J. *Designing and Conducting Health Surveys: A Comprehensive Guide*, 3rd ed.; 2006.

(106) The World Bank. World DataBank: World Development Indicators http://data.worldbank.org/indicator/SI.POV.NAHC/countries/GT?display=graph (accessed Jan 1, 2016).

(107) Central Intelligance Agency. Central America and Caribbean: Guatemala https://www.cia.gov/library/publications/the-world-factbook/geos/gt.html.

(108) Deaton, M.; Winebrake, J. J. *Dynamic Modeling of Environmental Systems*; Springer: New York, 2000.

(109) Hayduk, L. A. Shame for disrespecting evidence: the personal consequences of insufficient respect for structural equation model testing.

(110) Goffin, R. D. Assessing the adequacy of structural equation models: Golden rules and editorial policies. *Pers. Individ. Dif.* **2007**, *42* (5), 831–839.

(111) Hammond, R. a.; Dube, L. A systems science perspective and transdisciplinary models for food and nutrition security. *Proc. Natl. Acad. Sci.* **2012**, *109* (31), 12356–12363.

(112) Voth-Gaeddert, L.; Maxadoni Kamil, I.; Roosmini, D.; Oerther, D. Modeling Population Health Dynamics for Indonesia and Guatemala : An Application of System Dynamics Modeling. *Proc. 5th Environ. Technol. Manag. Conf.* **2015**,

VITA

Lee Voth-Gaeddert is a doctoral researcher in Environmental Engineering at Missouri S&T. He is a recipient of the NCAA postgraduate scholarship, a Schowalter Foundation grant, and the Chancellor's Fellowship at Missouri S&T. Mr. Voth-Gaeddert received his M.S. in Environmental Engineering and B.S. in Civil Engineering from S&T in 2014 and 2012, respectively. He received his PhD in Civil Engineering in July 2017 from Missouri University of Science and Technology.