What causes childhood stunting among children of San Vicente, Guatemala: Employing complimentary, system-analysis approaches

Lee E. Voth-Gaeddert[a],[*], Matthew Stoker[b], Devin Cornell[c], Daniel B. Oerther[a]

[a] Department of Civil, Architectural, and Environmental Engineering, Missouri University of Science and Technology, Rolla, MO, USA
[b] Independent Researcher, Quetzaltenango, Guatemala
[c] Department of Sociology, University of California, Santa Barbara, Santa Barbara, CA, USA

ABSTRACT

Guatemala has the sixth worst stunting rate with 48% of children under five years of age classified as stunted according to World Health Organization standards. This study utilizes two different yet complimentary system-analysis approaches to analyze correlations among environmental and demographic variables, environmental enteric dysfunction (EED), and child height-for-age (stunting metric) in Guatemala. Two descriptive models constructed around applicable environmental and demographic factors on child height-for-age and EED were analyzed using Network Analysis (NA) and Structural Equation Modeling (SEM). Data from two populations of children between the age of three months and five years were used. The first population (n = 2103) was drawn from the Food for Peace Baseline Survey conducted by the US Agency for International Development (USAID) in 2012, and the second population (n = 372) was drawn from an independent survey conducted by the San Vicente Health Center in 2016. The results from the NA of the height-for-age model confirmed pathogen exposure, nutrition, and prenatal health as important, and the results from the NA of the EED model confirmed water source, water treatment, and type of sanitation as important. The results from the SEM of the height-for-age model confirmed a statistically significant correlation among child height-for-age and child-mother interaction (-0.092, p = 0.076) while the SEM of the EED model confirmed the statistically significant correlation among EED and type of water treatment (-0.115, p = 0.013). Our approach supports important efforts to understand the complex set of factors associated with child stunting among communities sharing similarities with San Vicente.

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 et al., 2006; Dewey and 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 48% (International Food Policy Research Institute, 2016). Among rural Mayan communities in the western highlands of Guatemala, the rates of stunting are nearly 77% (ICF International, 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 et al., 2016). Millward (2017) and Mbuya and Humphrey (2016) provide recent, succinct literature reviews on the numerous hypothesized causal pathways of environmental factors impacting child stunting (Mbuya and Humphrey, 2016; Millward, 2017). Two factors of increasing concern for nutrient utilization within children are infectious disease exposure and fungal toxin exposure (Ali et al., 2016; Millward, 2017; Wild et al., 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). Furthermore, the dangerous combination high levels of maize consumption with high levels of aflatoxin presence in maize puts Guatemalan children at serious risk of high chronic exposure to this toxin (Immink and Alarcon, 1991; Torres et al., 2015). 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 and 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 (Voth-Gaeddert et al., 2015a,b). 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. Furthermore, we present a novel method to measuring EED, partially used by Kosek et al. (2017) by utilizing observable, symptom-based indicators and confirmatory factor analysis. The combination of these methods demonstrates a set of important tools 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 or aflatoxin exposure cause a reduction in child height-for-age?, (2) does prenatal health, child diet, or child-mother interactions cause a reduction in child height-for-age?, and (3) does poor water, sanitation, and hygiene practices cause an increase in 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 (ICF International, 2014) (n = 2103). 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 m, with an average range of temperatures of 5.1C–17C, and an annual rainfall of

1310 mm. Among the locations covered by the USAID survey, elevations varied between 1600 m and 3,100 m, temperatures varied between 9.5C and 20.9C, and annual rainfall varied between 800 and 2700 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)). The health assemblies included activities for all mothers and children regardless of health status or economic level, were promoted as a community-wide event, and were held at three different locations over three days of the same week to support equal participation among the various levels of health and economic statuses within the communities. 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 (ICF International, 2014). Institutional Review Board exemption for the use of de-identified data was attained from Missouri University of Science and Technology.

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 = 2103 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 and Cornell (2016) and the Python code can be found on Github (Cornell and Voth-Gaeddert, 2016).

The relationships identified in the spanning-tree graphics from the