Utilizing structural equation modeling as an evaluation tool for critical parameters of the biosand filter in a pilot study in Para, Brazil
L. E. Voth-Gaeddert, D. W. Divelbiss and D. B. Oerther

ABSTRACT

Biosand filters (BSFs) have been used to provide potable water to many communities throughout the world. A significant number of laboratory and field studies have demonstrated the effectiveness of the BSF if utilized properly. However, our prior work suggests that multiple factors contribute to the effectiveness of the BSF in a community setting. These factors include: household education levels (HELs), socio-economic status (SES), additional interventions for sanitation and hygiene, water sources, and the relationship between treatment and storage. A structural equation model (SEM) was constructed to evaluate the contribution of these factors to diarrheal occurrence in impoverished households in two Brazilian communities along the Amazon River. The results of this study showed that HEL was the most important factor in reducing diarrhea, and the presence of a BSF was near ineffective. Furthermore, interventions for sanitation and hygiene as well as SES all contributed to a reduction in diarrheal occurrence. This study demonstrates that SEM can provide a platform to evaluate the complex interaction among factors contributing to a reduction in the occurrence of diarrhea. Furthermore, the results of this study highlight the importance of a holistic approach to the deployment of technology-driven solutions such as the BSF.

Key words | biosand filters, Brazil, diarrhea, structural equation modeling

INTRODUCTION

The biosand filter (BSF) has been used for over 200 years (Centre for Affordable Water and Sanitation Technology 2012). A comprehensive body of literature has been established on the effectiveness of the BSF, including the model designed by the Center for Affordable Water and Sanitation Technology (Stauber et al. 2009; Tiwari et al. 2009; Vanderzwag et al. 2009; Aiken et al. 2011). Many of these studies show that in a controlled setting, whether in a laboratory or in the field, a removal of at least one log is achievable for a majority of harmful contaminants within water. However, this removal is highly dependent on the proper operation of the BSF. Recent reports suggest that the BSF has been deployed in more than 400,000 households in 60 different countries (Centre for Affordable Water and Sanitation Technology 2012). Across this spectrum of communities, development practitioners can expect to encounter variations in culture, education, wealth, and confounding health issues, as well as trends in demographic variables. In addition, fluids treated by the BSF are merely one of the ways in which diarrheal diseases may be communicated. Additional pathways for transmission include fingers, flies, floors and fields (World Bank 2014). To improve the understanding of this complex system and to identify critical, site-specific factors that are most cost-effective in reducing health issues, structural equation modeling (SEM) can be an effective tool.

Researcher-specified hypotheses relating components of a system under assessment are captured in the process of generating a SEM. The relationships among directly observable and hidden or latent variables can be captured in the model formulation. For example, socio-economic status (SES) is a type of latent variable that can be imagined
in the mind of the development professional, but cannot be directly measured. Instead, SES is inferred through measurements of directly observed characteristics including the choice of material for home construction, ownership of a boat, and the number of functioning light bulbs in a home. Additional features of SEM include the ability to perform simultaneous regression analysis, non-summation of errors within aggregated indicators, and a robust graphical interface. Using SEM, assessments of interventions to reduce diarrheal illness within communities have been reported previously for Guatemala (Divelbiss et al. 2013) and Brazil (Voth-Gaeddert et al. 2015). Latent and directly observable variables including filter operation and maintenance (FOM) of the BSFs, household education level (HEL), SES, diarrheal occurrences, household sanitation and water consumption habits were assessed simultaneously in a single SEM. The significance of understanding what factors have the largest effect within a community is crucial to a successful intervention.

**Background**

Divelbiss et al. (2013) studied communities located in the northern highlands in the Ixcan district of Guatemala. The study assessed the effectiveness of the BSF to reduce the occurrence of diarrhea. Variables considered in the construction of the SEM included the HEL, SES, source of water, additional water treatment and use of soap within the house. The primary hypothesis was that the BSF was the most significant factor in reducing diarrheal occurrences; however, the results of the study failed to support this hypothesis. Rather, Divelbiss et al. (2013) identified HEL as the primary contributing factor, followed by improved water source and then FOM. Given the unexpected nature of these results, further investigation was warranted.

A secondary study site was selected in the region of Para, Brazil to allow the re-development of the SEM. A partnership was formed with a local non-governmental organization (NGO) that had distributed over 10,000 BSFs among rural households throughout the region. In an initial feasibility study, Voth-Gaeddert et al. (2015) successfully demonstrated that the SEM that was developed to describe communities in Guatemala could be adapted to describe communities in Brazil. Although the feasibility was limited, the preliminary results of Voth-Gaeddert et al. (2015) confirmed the prior conclusion of Divelbiss et al. (2013), namely that the BSF, while important, was not the most significant factor responsible for the reduction in diarrheal occurrence observed in the communities. Improved water source and education were found to be the most significant factors in reducing diarrheal occurrences in the feasibility study.

This paper reports follow-up work at the pilot-scale to confirm the feasibility study and to employ an updated SEM and a household survey to gather and analyze a more substantial data set. For this expanded pilot-study, the central hypothesis remained the same, namely, that the BSF has the largest negative effect on diarrheal occurrences in households in Para, Brazil.

**METHODOLOGY**

**Location**

Through expert advice from the local NGO, two villages were selected for collaboration, namely Carapanatuba and Cabeça Donca (02°06′18.0″S 54°45′30.0″W and 02°05′46.9″S 54°45′15.0″W). Both villages are located in the same municipality, Santarem, Para, Brazil, and this is the same municipality evaluated in the prior feasibility study (Voth-Gaeddert et al. 2015). Through interviews with local leaders, fishing and crafting of souvenirs were identified as the primary means of employment for males and females, respectively. Extreme poverty (that is, income less than $1.25 per person per day) was not common; however, not all basic needs were satisfied within households. Houses were constructed on stilts to accommodate seasonal flooding of the Amazon River. Depending on the location of the village, flooding occurred during 3–5 months of the year. During the flood, canoes were used for door-to-door transportation. The flooding created problems in several areas, but especially in the transmission of fecal material. While pit latrines were commonly used during the dry season, when flood waters were present, simple latrines were mounted on the back of each house and fecal material was simply dropped into the flood water. As the drinking water was collected directly from the river, the water...
source was substantially contaminated with fecal material during the flood.

Data collection

Interviews were conducted through a translator. During the initial interviews with local leaders, ‘poor’, ‘average’ and ‘wealthy’ categories were identified, and the team randomly selected a total of 41 households that included balanced representation among the three categories. For Carapanatuba, 17 homes were sampled, and for Cabeca D’onca a total of 24 homes were sampled. Interviews consisted of a survey delivered verbally, and a visual inspection of the BSF in the home. The survey was modified from the survey used in the prior feasibility study (Voth-Gaeddert et al. 2015) based upon field observations. A total of 49 questions were part of the verbally administered survey (see supplementary material, available online at http://www.iwaponline.com/ws/015/041.pdf). The following modifications were made due to field observations.

Throughout this section, variables are listed first, and then an explanation follows:

- HEL observable variable ‘Reading and Writing’ test was administered

The previous SEM model (Voth-Gaeddert et al. 2015) was only able to use two observable variables to describe HEL where the literature recommends three or more. A reading and writing test was added because it did not allow the interviewee to hide illiteracy and helped to more accurately describe cognitive ability of the household.

- SES observable variables of ownership of a TV, shower, working light bulbs, material of water storage container, and ‘improved’ roof, along with others, were collected

A wide spectrum of SES indicators can be found in the literature including: income level, type of job, net worth of physical objects found in the household, type of objects in household or a combination of these (Dressler et al. 1998; Jones et al. 2011). Further study is needed within this particular subject on strength of indicators, as this can vary widely between sub-regions of a country. Physical objects found in the household were the main indicators for SES within this study.

- Diarrheal health burden (DHB) was changed to diarrheal occurrences, which included diarrhea most prevalent within the house and if a household member had diarrhea within the past 2 weeks

This transition from DHB to diarrheal occurrences was due to a lack of available indicator variables. The low occurrence of diarrhea during the dry season of the Amazon River limited the sensitivity of the latent variable; however, an observable endogenous variable proved robust. During the rainy season, the Amazon River floods the latrines and carries the contents downstream past other villagers’ households.

- Independent variables were changed: ‘Soap & Towel Present’ to ‘Improved Sanitation’ and definition of terms changed

While the presence of soap and towel within a household remains important, it was found that over 90% of the sample population had these items. This term was then incorporated into the ‘Improved Sanitation’ score partly defined by the World Health Organization (WHO 2006). After further investigation, the variable ‘Improved Storage’ was added to test the hypothesis that if a household stores their water properly, the occurrence of diarrhea within the household will be reduced. In addition, ‘Adequate Sanitation’ was changed to ‘Improved Sanitation’ to be consistent with WHO terminology (WHO 2006).

Data analysis

While there are multiple ways to utilize SEM, this study develops an exploratory technique. An original hypothesized SEM (i.e. Voth-Gaeddert et al. (2015), before data collection) is adjusted to ‘match’ the empirical data acquired from a sub-population (i.e. Voth-Gaeddert et al. (2015), after data collection). Once an adjustment is made, the procedure becomes exploratory and needs to be confirmed with a new set of data. Confirmatory factor analysis (CFA), as the name suggests, is a confirmatory technique and is used to confirm changes to the original SEM. When CFA is used, one hypothesized model (outcome of Voth-Gaeddert et al. 2015) is being compared with a data-driven model (data from altered survey of the two new locations) using the covariance matrices of both for analysis. A minimization
of the difference between matrices is the goal of CFA and confirmed through test of model fit. These tests include chi-square ($p > 0.05$; goal is to not reject the null), root mean square error of approximation (RMSEA < 0.10), comparative fit index (CFI > 0.90) and Tucker Lewis index (TLI > 0.90; Schreiber et al. 2006). The software package Mplus 7.2 was used for the analysis of data (Muthen & Muthen 1998–2012). The software manufacturer recommends that the model pass three of the four tests to proceed. If the model fails more than one test, adjustments are made to the SEM; these adjustments are dictated by several output statistics offered by Mplus 7.2, including modification indices and correlated residuals. As adjustments should not be based on statistics alone, the literature should be consulted to confirm alterations. Each iteration drives the model toward a better representation of reality within the region (Grace 2006).

A full SEM incorporates latent variables, depicted by circles, observable variables, depicted by squares, and arrows, depicting hypothesized causal relationships. Each latent variable has three or more observable indicator variables that are ‘reflected’ or ‘manifested’ in the latent variable. An analysis of only the latent variables and their indicator variables is created to test the acceptability of the representation of the indicator variables (see Figure 1). This is referred to as the ‘measurement model’ and is used to verify that the latent variables are represented correctly. Once tests of model fit confirm the measurement model is correct, ‘path analysis’ (simultaneous regression analysis) is used to assess factors acting directly or indirectly on the diarrheal occurrence variable. This is termed the ‘structural model’.

Once the full model has passed three of the four tests of model fit, the path coefficients can be assessed. Path coefficients are given in standardized and unstandardized results. Unstandardized variables are informative when ordinal response units add value to the answer. As WHO guidelines were used in scoring many of the variables, standardized path coefficients are adequate and allow cross-comparison between variables. The path coefficients can be interpreted as standard deviation unit increases between variables. For example, a direct path from SES to diarrheal occurrences with a value of $-0.071$ means that as SES increases by one standard deviation unit (an increase in ‘wealth’), diarrhea will decrease by $0.071$ standard deviation units (improved health). However, it is stressed within the literature that causal assumptions are held for the reader to determine but can be addressed within the discussion section of a report (Kline 2005).

A continued awareness and analysis of cultural differences is crucial when working with multiple global locations. For this reason, changes to the previously reported survey and SEM model (Voth-Gaeddert et al. 2015) were proposed based upon a deeper understanding of cultural influences. These changes occurred mainly within the observable indicator variables used to represent the latent variables in the model. All changes were made either before data collection began, or at the time of transition from confirmatory analysis to exploratory analysis. Understanding impacts of BSFs in communities around the world warrants a continued devotion to cultural observations.

**Limitations**

Local leaders were used to help identify different levels of SES households to improve the spectrum of representation.
Randomization was then used to select the individual households, hence clustered.

Seasonal flooding occurred in both villages; during data collection the majority of households were not flooded. While the fluctuation of diarrheal disease transmission pathway loads (or the dynamic average probability of getting sick through any of the pathways at one time) has not been studied for Brazil, all transmission pathways remain relevant regardless of seasonal flooding variability. Therefore, results were not affected.

The sample size of this study was 41 houses collected as part of the second iteration. Iacobucci (2010) offers a minimum suggested sample size for SEM of 50, whereas other authors offer larger recommendations (Ullman 2006; Schreiber et al. 2006; Barrett 2007). A small sample size can result in increased measurement error, which will cause the over-estimation of the chi-square value (meaning it will identify a significant difference between the data-driven model and the hypothesized model, failing to confirm a good fitting model). However, all model fit tests, including the chi-square, reported a robust fit, indicating that the small sample size in this study was sufficient. In addition, the model was identified and multicollinearity was not an issue.

**RESULTS AND DISCUSSION**

From the revised survey and SEM, a new theoretical model was created and data from 41 households were used to populate the altered hypothesized model. A measurement model was used to assess the fit of observable indicator variables on their respective latent variables. From the four tests of model fit, only one was satisfactory ($\chi^2 p > 0.002$; RMSEA = 0.340; CFI = 0.72; TLI = 0.75), meaning at least one of the chosen indicator variables did not accurately reflect the latent variable. After investigation of the indicator variables, ‘Improved Roof’ was identified as the variable that did not depict the SES latent variable accurately. As alterations to a SEM model should not be data driven alone, alternative models were assessed based on discussion with practitioners and additional review of the literature. Due to the absence of literature on improved roofs in Brazil as well as recommendations from local practitioners, the variable was removed and the measurement model was revaluated. All four tests of model fit confirmed the use of the selected indicator variables ($\chi^2 p > 0.284$; RMSEA = 0.073; CFI = 0.98; TLI = 0.97). Once the measurement model was confirmed, the full model could be evaluated.

Within the SEM methodology, the same four tests of model fit are used for the full model. The full model in this study showed that the data accurately explained the hypothesized relationships ($\chi^2 p > 0.617$; RMSEA = 0.000; CFI = 1.00; TLI = 1.08). Figure 2 depicts the full model that graphically depicts all of the hypothesized relationships. Figure 2 also provides the computed path coefficients. HEL had the highest significant negative effect on the diarrheal occurrences (−0.278). This was followed by improved sanitation and SES (−0.200 and −0.071, respectively). Positive relationships affecting FOM included improved water source, improved sanitation and HEL (0.353, 0.277 and 0.256, respectively). The only prohibiting variable on FOM was additional water treatment (−0.344). The explained variance for FOM and diarrhea was 41% and 28%, respectively.

The data showed that HEL was the most important variable in reducing diarrheal occurrences within the two villages. As households became more educated, their understanding of the diarrheal disease transmission pathways also increased, helping to block all pathways, not just the fluids pathway. Similarly, a family who understood how to put the knowledge or education to use through improving sanitation practices was also able to block multiple pathways. Families who spent time trying to change water sources and improve sanitation actually harmed themselves with regards to diarrhea. Some parts of the villages had access to a piped network, but further investigation revealed that the water still came from the Amazon River. The issue of water storage could have arisen from the extra step that storing the water produces within the use of the BSF, increasing the opportunity for the water to become re-contaminated.

The FOM was negatively affected by the use of additional water treatment techniques. Many of the villagers possessed resources to acquire easier treatments, such as chlorine. If the difference in protection from diarrhea was not significant, the less time-intensive treatment was used. Both improved water source and improved sanitation positively influenced the FOM. One explanation is that if the household is going to make the effort to improve other sanitation and hygiene-related practices, an effort to improve the
use of the BSF will also occur. HEL also improved the FOM. As households increased in overall education, the understanding of how the BSF operates also increased. Finally, the FOM had a near-negligible effect on diarrheal occurrences. This is envisioned to be caused by the high SES of the population. From the results, it appears that households had the money to reduce transmission in ways that were less intensive and more effective.

CONCLUSIONS

While this study successfully identified critical diarrheal disease transmission pathways, the applicability of the pilot result is limited to two villages in Para, Brazil. Further study is needed if an understanding of the broader region is to be established. However, within the villages studied, HEL is the most significant factor for reducing diarrhea occurrence. This is contrary to the original hypothesis that the BSF is the most important factor in the reduction of diarrhea. Improved sanitation practices and SES also reduced diarrheal occurrence within the households. These results are significant for two reasons. First, if organizations are to work within this region, this study provides a base of understanding the complex interrelations among environmental factors and interventions for reducing diarrheal transmission. Second, this study highlights the importance of recognizing the potential for a technology-driven solution to under-deliver when placed in the complex context of a field situation; therefore a holistic assessment is recommended to evaluate the most effective intervention to aid a specific, target population.

REFERENCES


First received 31 October 2014; accepted in revised form 19 March 2015. Available online 1 April 2015