

## Clean Water for Liberia

Kate McLain

Professor Pruim and Professor De Ruiter

Sawyer Products and Liberian NGOs are partnering with students and faculty from Calvin College in statistics, geography, and public health for a four-year project studying the efficacy of distributing water filters to households in Liberia. The overall goal of this project is to provide clean drinking water to approximately 100,000 households in Liberia. At the time of distribution, Liberian NGOs are conducting surveys with basic questions on demographic information and health-related issues. More data are collected at follow-up visits about two and eight weeks after the installation of the water filters to determine their efficacy.

The role of our team in the Mathematics and Statistics Department was to clean and reshape the data, provide analyses for various elements of the data, and fit statistical models. Our team received regular updates of the data from the NGOs in Liberia conducting the surveys on tablet computers. For the data cleanup and analysis, our team used R, a programming language used for statistical computing.

This summer was the first summer of the project and therefore entailed more data cleaning than analysis. While the survey seemed straightforward, there was much room for error and discrepancies in the recording of data. I helped to clean the data and prepare it for analysis. Our team also brain-stormed about different variables that would be interesting and useful to compare in this study. We collaborated with the other teams from Calvin and provided them with data, graphics, and summary statistics useful for their aspects of the research. Through the process of data cleaning, our team was able to update Sawyer on issues with the data collection and survey form and assist them in improving the collection process and tightening up the survey.

Since this was the first summer of a longer-term project, there are not any final results to report yet. However, in our initial analysis, we did find a decrease in the symptoms from the initial to the follow-up visits, which could be related to the use of the filters. Diarrhea is a common symptom related to poor water sources. We were able to fit some basic statistical models and found that the filters were effective in reducing the odds of diarrhea in children under five by ninety-five percent.

Coming into this project, I had minimal experience working in R and coding. While it was overwhelming at first, I made significant gains in these areas over just a short period of time and learned a lot about working independently and teaching myself new things. This project gave me a taste for data analysis and working with a large, complex data set. Although the messiness of the data set was frustrating, it gave me experience working with a real-world data. Additionally, I gained experience working with an actual company and providing quality summaries and reports for their use. I am very grateful for this experience and the role I could play in this larger project to provide clean water to those who do not have access.