Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Gouthami Rao Date

Microscale Dynamics of *Escherichia coli* in Rivers of Northern Coastal Ecuador

By

Gouthami Rao
Master of Public Health

Environmental Health

_____

Karen Levy, PhD, MPH
Committee Chair

_____

Paige Tolbert, PhD
Committee Member

Microscale Dynamics of *Escherichia coli* in Rivers of Northern Coastal Ecuador

By

Gouthami Rao

Bachelor of Science
Emory University
2012

Thesis Committee Chair: Karen Levy, PhD, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Environmental Health
2013

Abstract

Microscale Dynamics of *Escherichia coli* in Rivers of Northern Coastal Ecuador

By Gouthami Rao

Background: Diarrheal disease is a leading cause of mortality and morbidity around the world, killing approximately 1.5 million children and resulting in 2 billion cases every year. The use of contaminated water continues to be a pressing issue, with approximately 1.1 billion people who lack access to improved water sources in the world. In northern coastal Ecuador many communities rely heavily on untreated surface water as their primary source of drinking water. These villages routinely access the river for water for drinking, cooking, bathing, washing, navigation, and recreation, and through these interactions the streams become contaminated.

Methods: We undertook a study to explore how specific water collection locations and microscale river hydrodynamics affect microbial water quality on three rivers with varying stream velocity and turbidity profiles (Rio Santiago, Rio Onzole, Rio Cayapas). We carried out this research in six villages in the Esmeraldas Province of northern coastal Ecuador during June-July, 2012. Our study focused on the following questions: (1) How does distance from river shore affect microbial contamination levels; and (2) How do physicochemical water quality variables affect microbial contamination levels in the river. We collected a total of 355 water samples and tested for *E. coli* concentrations using the IDEXX Quantitray method. We established 2-3 transects in each village at sites where villagers washed clothes and dishes, and collected water. Each transect consisted of six point samples to cover locations both within and outside of the river eddy.

Results: We found a significant association between proximity to shore and *E. coli* concentrations. Higher *E. coli* concentrations were also significantly associated with increased turbidity and decreased dissolved oxygen levels.

Conclusion: The results of this study can help inform community members about the safest locations to collect drinking water and can also provide information to characterize and differentiate rivers based on localized contamination of microbial contaminants and water quality parameters.

Abstract

Dinámicas a microscala de *Escherichia coli* en ríos de la costa norte del Ecuador

Por Gouthami Rao

Antecedentes: Las enfermedades diarreicas constituyen una causa importante de mortalidad y morbilidad a nivel mundial, causando la muerte de aproximadamente 1,5 millones de niños y dando lugar a 2 mil millones de casos al año. El uso de agua contaminada sigue siendo un problema apremiante, con aproximadamente 1,1 millones de personas que carecen de acceso a fuentes de agua potable mejorada en el mundo. Al norte de la costa del Ecuador muchas comunidades dependen en gran medida del agua superficial no tratada como su principal fuente de agua. La gente de estas comunidades accede habitualmente el río para recolectar agua para beber, cocinar, bañarse, lavar, navegar y recrearse, y a través de estas interacciones se contaminan los arroyos.

Métodos: Se realizó un estudio para explorar cómo localizaciones específicas de recolección de agua y la hidrodinámica a microescala de los ríos afectan a la calidad microbiológica del agua en tres ríos con variaciones en la velocidad de la corriente y perfiles de turbidez (Río Santiago, Río Onzole, Río Cayapas).

Esta investigación se llevó a cabo en seis comunidades de la provincia de Esmeraldas en la costa norte de Ecuador durante junio-julio de 2012. El estudio se enfocó en las siguientes preguntas: (1) ¿Cómo la distancia a la orilla del río influye en los niveles de contaminación microbiana? y (2) ¿Cómo las variables físico-químicas de calidad de agua afectan a los niveles de contaminación microbiana en el río? Se recolectó un total de 355 muestras de agua, a las cuales se les realizaron pruebas de concentración de *E. coli* con el método IDEXX Quantitray. Se establecieron 2-3 transectos en cada comunidad en los sitios donde los pobladores lavan la ropa y los platos, y recogen el agua. Cada transecto consistió en seis muestras puntuales para cubrir lugares tanto dentro como fuera del remolino del río.

Resultados: Se encontró una asociación significativa entre la proximidad a la orilla y las concentraciones de *E. coli*. Las elevadas concentraciones de *E. coli* también tuvieron una asociación significativa con el aumento de la turbidez y la disminución de los niveles de oxígeno disuelto.

Conclusión: Los resultados de este estudio pueden ayudar a informar a los miembros de las comunidades sobre los lugares más seguros para recoger agua y también pueden proporcionar información para caracterizar y diferenciar los ríos en base a la contaminación bacteriana localizada y los parámetros de calidad del agua.

Microscale Dynamics of *Escherichia coli* in Rivers of Northern Coastal Ecuador

By

Gouthami Rao

Bachelor of Science
Emory University
2012

Thesis Committee Chair: Karen Levy, PhD, MPH

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Environmental Health
2013

Acknowledgements

I would like to first acknowledge Dr. Karen Levy for her engaged guidance and committed mentorship these past three years. I would like to acknowledge the endless support from Gabriel Trueba, William Cevallos, Chelo Ortiz and Sixto for their commitment to this project and its execution in Ecuador. Also, I am grateful for the advice and guidance on the analysis from Baiyu Yang, Dr. David Kleinbaum, William Oswald, Marissa Grossman, and Sharia Ahmed. Lastly, I am incredibly thankful for my friends and family who have continually supported me in my passion and dreams to work in the field of environmental health.

Table of Contents

INTRODUCTION

Diarrheal Diseases and Water use

Diarrhea is the fifth leading cause of mortality and third largest cause of morbidity around the world, killing approximately 1.5 million children and resulting in 2 billion cases every year (WHO 2009). Diarrheal diseases are especially persistent in developing countries and primarily affect children under 5 years of age. Diarrhea is known to be a symptom of an intestinal tract infection from bacterial, viral, and/or parasitic microorganisms. Infection is primarily spread through feces-contaminated food, drinking water, or person-to-person (WHO 2009). The use of contaminated water continues to be a pressing issue in developing countries, with approximately 1.1 billion people who lack access to improved water sources in the world (WHO 2000).

Issues of water scarcity

A lack of overall global water resources perpetuates the struggle for clean water sources in developing countries. Water security is linked to water availability, and only approximately 25% or 12,000 km^3/year of the global total river runoff and groundwater recharge is available for human use (after subtracting uncaptured storm runoff), and nearly 40% (or 5000 km^3/year) of this fraction is already being withdrawn for human use from rivers, lakes and groundwater (Heathwaite 2010). Ten percent of the population in the least developed countries, as designated by the United Nations, relies on surface water, which is considered an unimproved water source for drinking water. Surface water is defined as water collected directly from rivers, lakes, ponds, irrigation channels and other surface sources (UNICEF and WHO 2012). Rivers are the source of most of the

fresh surface water used by humans, but they only constitute about 1/10,000th of one percent of the total global fresh water (USGS 2012).

In developing countries, the limited amount of freshwater resources force people to rely on available sources. The world's six billion people are appropriating 54% of all the accessible freshwater contained in rivers, lakes and underground aquifers (WWAP). Currently, 900 million people rely on unimproved drinking-water supplies where this statistic continues to grow and until further progress occurs, developing countries will inevitably be faced with consuming contaminated drinking waters (WHO/UN-Water 2008).

Recontamination of water sources and transmission processes of diarrheal diseases

Particularly in developing countries, the limited resources of freshwater have forced many to use, reuse, and likely re-contaminate, water sources through various anthropogenic activities. In Northern Ecuador, cleaning cloth diapers, washing dishes, and bathing are all common practices that inevitably impact and potentially contaminate the local rivers. Where water resources are limited villagers often rely on river water to be multi-purposeful. This occurrence is not limited to observations in Ecuador, but domestic use of source water occurs in other countries as well such as Bangladesh (Begum, Talukder et al. 2005), India (Hamner, Tripathi et al. 2006), Nigeria (Strauch and Almedom 2011), and others.

Transmission processes influenced by human behavior play an important role in contracting diarrheal diseases. Fecal indicator bacteria (FIB) are a long-standing measurement tool used to determine the sanitary quality of water (Edberg, Rice et al.

2000; Ishii and Sadowsky 2008). Instead of detecting specific pathogens, indicator bacteria, such as *Escherichia coli*, are generally used as a measurement tool to determine recent fecal contamination (Edberg, Rice et al. 2000; Levy, Nelson et al. 2012). *E. coli* is excreted from all warm-blooded animals and some reptiles (Ishii and Sadowsky 2008; Lyautey, Lu et al. 2010). An increase in levels of fecal contamination indicates areas where the potential for transmission of diarrheal disease is more likely to occur.

Spatial variability of indicator bacteria

Indicator bacteria are used to detect levels of fecal contamination in water, but little is known about the spatial variability of such bacteria. Several variables contribute, temporally and spatially, to the quantity of fecal indicator bacteria in surface waters. Known temporal variables include rainfall, seasonal variations, and runoff, but less is understood about spatial variability in surface waters. Several studies have shown that localized re-suspension along the banks allows for *E. coli* and other bacteria to persist in the sediment. It has been particularly noted that persistence and replication occur readily in tropical and subtropical soils and water (Quilliam, Clements et al. 2011), (Byappanahalli, Fowler et al. 2003), (Fries, Characklis et al. 2008). Between the banks and the central flow of rivers, the die-off of indicator bacteria caused by spatial variability is still unexplained (Hellweger and Masopust 2008).

Spatial variability in a site relates to the sources within a site, advection, distribution of mixing at the site, and occurs at various scales and directions (EPA 2010). Two competing ideas to explain the microbial spatial variability in rivers include (1) bacterial loads might increase downstream or (2) die-off occurs. Several studies

comparing upstream and downstream samples indicated an increase in fecal contamination further downstream (Byappanahalli, Fowler et al. 2003; Quilliam, Clements et al. 2011). On the other hand, die-off occurs with increasing time and distance from the point of contamination (Quilliam, Clements et al. 2011). Although, exact points of contamination can be hard to detect, the transport and fate of *E. coli* as indicator bacteria can produce a general idea of contamination levels in these surface waters.

Purpose of study

Water quality and quantity continue to be an issue in developing countries, but a stronger understanding of the links between levels of contamination and human contact with and ingestion of water can lead to improved forms of intervention methods. By understanding the spatial variability of microbial contamination there is a possibility of creating a village-level intervention to collect water in recommended areas. If contamination occurs on the riverbanks, and often in eddies, then contamination should decrease with increasing distance from shore and moving towards the central flow of a river. Also, there is interest in die-off of bacteria between villages to better characterize the rivers based on *E. coli* concentration patterns per village. This study explores these issues through distance from shore and turbidity flow rates to microscale dynamics for three different rivers of the Esmeraldas Province in northern coastal Ecuador.

In northern coastal Ecuador, the EcoDess project investigates diarrheal disease transmission in 24 communities that struggle with issues of diarrheal disease in the Esmeraldas province. The current knowledge about water quality in the Northern Ecuador area, studied by EcoDess, remains that rivers serve as the primary water source

for 68% of households in the region, and 60% of households report drinking their water without treating it. This study area has been observed to have high rates of diarrheal disease (Levy, Hubbard et al. 2009; Levy, Nelson et al. 2012).

The specific objectives of this study were:

1) To understand how distance from river shore affects fecal contamination levels.
2) To characterize and differentiate the study rivers based on water quality parameters and levels of localized fecal contamination.

Understanding the microscale dynamics of rivers that are often used as primary source water is important in determining potential areas for disease transmission and developing public health intervention recommendations.

METHODS

Study Region

Several of the 24 communities involved with EcoDess rely heavily on untreated surface water from the river as their primary source of water. These river villages routinely access the river to use the water for consumption, drinking, cooking, bathing, and recreation, and through these interactions the rivers become heavily contaminated. Field activities were conducted in summer 2012 in partnership with and supplemental to EcoDess' research activities.

Samples were collected over 24 sampling days, including one pilot (June 5-6, 2012) and three field visit periods (June 14- 21, June 24- July 1, and July 11-19). All villages selected for the study were accessible by boat and primarily utilized the respective river for washing clothes, dishes, and collecting water. The pilot period took

place in a village called Rocafuerte. Visit 1 consisted of two river villages: La Peña and the completion of the Rocafuerte pilot period. Visit 2 consisted of two other river villages: Arenales and Tangare. Lastly, Visit 3 included sampling from Telembi and Trinidad. Total numbers of samples analyzed per village can be seen in Table 1.

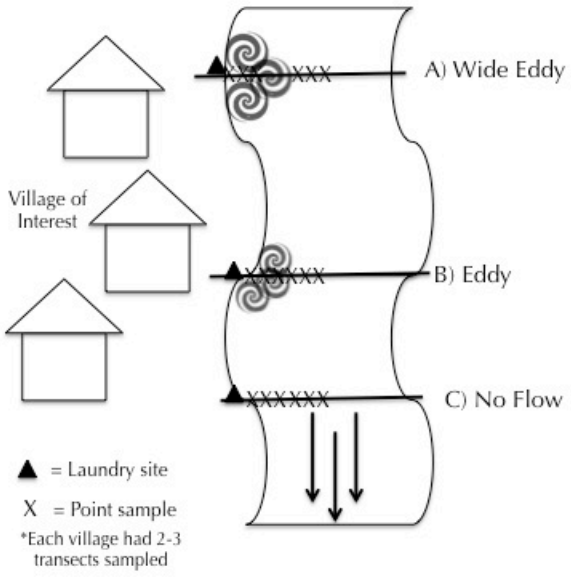
Table 1. Total number of samples per village by river

Village	River	Total samples (n)
Rocafuerte	Santiago	30
La Peña	Santiago	72
Arenales	Onzole	48
Tangare	Onzole	72
Telembi	Cayapas	48
Trinidad	Cayapas	72

Study Design

The primary goal of this study was to investigate the association between *E. coli* and distance from shore. River water samples were collected for analysis in areas of rural villages with frequent human-river interaction. Secondarily, river water samples were used to characterize and differentiate different types of rivers based on physicochemical water quality variables. A further description of the field sampling schematic is seen in Figure 1 and described in sampling methods.

Figure 1: Field sampling schematic. A) All samples collected 2m apart except the third and fourth samples as the break-point for samples 'within eddy' and 'outside eddy' B) & C) All samples collected 2m apart.



Sampling Methods

Six river villages along three rivers were studied in northern coastal Ecuador. Two villages were selected per river: La Peña and Rocafuerte for Santiago, Arenales and Tangare for Onzole, and Trinidad and Telembi for Cayapas. The pilot period methods are seen under "Rocafuerte (pilot)," and sampling collection methods were specific to each river (see: River-specific sampling collection). Each village had between 2-3 transects and were selected based on local villagers who indicated where they wash clothes and dishes. These areas were designated as the laundry site (Figure 1). At each site, a transect was established perpendicular to the shore, and six point samples were collected along each transect. In addition to water sample collection, measurements were taken at each point for pH, turbidity (NTU), temperature (°C), dissolved oxygen (DO) (ppm), and stream velocity (ft./sec). Details on the measurement of these parameters is provided

below. The majority of sites were located at the base of community stairways. All samples were collected in a canoe lent by a villager or the EcoDess boat. A marker, typically the lower section of a distinctive tree or bush, was chosen directly across the river from the laundry site to determine the total river width using a Yardage Pro range-finder (Bushnell, Overland Park, KS) tool. Samples were then collected every two meters, by subtracting the total river width from the current river width. The current river width (i.e. distance away from the marker) was validated every time a point sample was collected, by checking the range finder's value 3-4 times. A daily diagram of the river activity at sampling time for all transect samples was drawn including length of river at transect, locations of sampling points, water activity, and other significant markers. A simplified, generalized version of these diagrams is shown in Figure 1.

All village water samples were collected between 10:00-11:00 AM, taken to the microbiology field lab immediately, stored on ice, and processed within 6-8 hours of collection. A negative control sample was processed every day using deionized water (n=24).

Rocafuerte (Pilot):

On the first day of the pilot period conducted in Rocafuerte, two transects were selected based on the two areas where local villagers washed clothes and dishes. In order to determine if these river water samples required dilutions, one transect of samples was processed with a 1:10 dilution and the second transect without dilutions. For transect one, three samples within the eddy and three samples outside the eddy were taken at the

central flow of the stream (>70m from shore). For transect two, all samples were consecutively taken approximately 1.5 meters apart. On the second day of the pilot period, transect one included three samples within the eddy and the three samples outside the eddy. Although, on this day the samples collected outside the eddy were closer to shore because it was unreasonable to assume people would collect water >70 m from shore (compared to day one). On this day none of the samples from either transects were diluted.

Santiago River (fast-flowing):

Sampling from the Santiago River began with the pilot period, continued with La Peña, and completed the remaining two days in Rocafuerte. Due to the strong river currents and rocky bottom riverbed, three anchors were used to stabilize the boat: one in the front, middle and back. The assistant navigated the boat to the site of interest and the range finder was used to measure the full width of the river. Once the width was recorded, the assistant maintained the boat's parallel position to the shore while the researcher collected a point sample from the surface water with a 100 mL sterile Whirl-Pak bag (Nasco, Fort Atkinson, WI). Immediately after sampling, the DO, pH, temperature, turbidity and stream velocity measurements were taken at the same point. If a boat passed by, then measurement sampling was halted for 5 minutes while the river water returned to its normal flow. The next point sample was taken by subtracting the previous point sample river width to the current river width, and if the difference was 2 feet then the anchor was readjusted and sampling and measurements ensued. Eddy lines were easily distinguishable between the no flow and high flow areas. For wide eddies,

three point samples were taken within the eddy line and three samples taken outside of the eddy line.

Onzole River (slow-flowing):

Sampling from the Onzole River required minimal field expertise for this river. The slow flow of the river and dirt riverbed only necessitated one anchor to maintain the boat's position by the assistant. All sampling procedures to record river width, collect samples, and measure readings were the same as the Santiago river.

If the eddy size was regular, then three point samples were taken 2 meters apart within the eddy line. The remaining samples were taken right outside of the eddy line with a 2 meters distance between each sample. If there was no flow (i.e. no distinguishable eddy line) then all six point samples were taken 2 meters apart. If the eddy size was wide, similarly to Santiago, then three point samples were taken within the eddy line and the remaining three samples outside of the distinguished eddy line, even if that included a gap between point samples #3 and #4.

Cayapas River (intermediate-flow):

Sampling from the Cayapas river used similar techniques to that of Santiago and Onzole. The assistant used two anchors: one in the front and one in the back to maintain boat position.

Measuring Probable Fecal Contamination at river sites

E. coli samples were collected with Whirl-Pak bags and enumerated using IDEXX quanti-trays (IDEXX, Westbrook, ME). The Whirl-Pak bags were submerged under water to a typical depth of water collection by community members, according to the field assistant, and filled up to the 100 mL line marked on the bags.

Measuring Stream Velocity

All instruments used were calibrated before each visit. Instantaneous velocity was measured with the Flow Probe (Global Water Instrumentation Inc., Model FP111, College Station, TX) to measure subsurface velocity of point samples. The probe was placed at a depth of typical water collection and held for 15 seconds.

Measuring pH

Temperature and pH were measured using Hanna Instruments waterproof pH tester (HANNA, Woonsocket, RI). De-ionized water was poured into a sterilized Whirl-Pak bag. Then, the pH meter was turned on and put it into the de-ionized water to calibrate. To take each sample, the electrode was rinsed with de-ionized water. Then, the pH reading stabilized after 5 seconds. The electrode was rinsed after readings for all samples collected on a given day.

Measuring dissolved oxygen

First, the YSI dissolved oxygen (DO) handheld probe (YSI Inc., Yellow Springs, OH) was calibrated. The probe membrane was rinsed with distilled water. Using a sterile Whirl-pak, a zero-oxygen water sample was created to calibrate the dissolved oxygen meter. The probe was then submerged into the Whirl-pak bag deep enough to cover the membrane and automatic temperature compensation element. Then the dissolved oxygen reading was recorded. Next, the probe was rinsed with distilled water and capped. For each DO reading, the probe was rinsed with distilled water after taking the readings for the day. Since the DO probe is very sensitive, it was calibrated before taking measurements at each transect.

Measuring turbidity

Using the Hach 2100Q Turbidimeter (HACH Chemical Company, Loveland, CO), the meter was first calibrated using the standards given before entering the field for each visit. A subsurface river water sample was collected in a clean vial. The vial was filled and rinsed three times before taking the final sample. Then the vial was dried and inserted into the meter for measurement.

Laboratory Methods

All river water samples were kept on ice and processed using IDEXX methods within 24 hours of collection. One Colilert reagent packet was added to each Whirl-Pak bag and was vigorously shaken. When the powder reagent completely dissolved, the sample was poured into a Quanti-Tray, and securely sealed using the IDEXX Quanti-Tray sealer. When voltage was too low (<220V) for the sealer to turn on, a conventional

iron was used to seal the trays, ensuring that all wells contained sample water. This only occurred for one day of samples on June 14th, 2012. The processing date and sample ID was marked on the back of trays and also in the lab data sheet. All 100 mL samples were processed straight from Whirl-pak bags. If turbidity levels were high then a 1:10 dilution was performed using sterile syringes to extract 10 mL of the river water sample. Syringes were boiled and placed in a zip-lock bag with alcohol-soaked paper towels to avoid risk of environmental contamination. Dilutions were used with sterile DI water. Sterile DI water was prepared and autoclaved before entering the field conditions, but when sterile DI water was not available bottled water was used as a replacement.

Trays were incubated for 18-24 hours at 41+/- 3 degrees Fahrenheit in a scientific incubator (Boekle, Feasterville, PA). A voltage regulator was used to maintain voltage as best as possible. When the village had energy cuts a generator was used to maintain power because the IDEXX sealer required 220V at all times. Results from the microbial analysis were read in Most Probable Number (MPN) of Colony Forming Units (CFUs) for *E. coli*.

Data Analysis Methods

Data entry and cleaning

All hand-recorded field data sheets were entered into one spreadsheet in Microsoft Excel for data cleaning and statistical analyses. In order to ensure accuracy, 10% of data was entered twice into Microsoft Excel. Quality assurance of entry yielded a 0% error rate and the data was therefore considered accurate. Data analysis was conducted using SAS v9.3 (Cary, NC), and graphics were produced in STATA v12 (College Station, TX).

A total of 332/355 (93.5%) samples fell within a countable *E. coli* range. A total of 12/355 (3.4%) samples were above the detection limit and treated as the maximum countable 2419.6 MPN. A total of 11/355 (3.1%) were under the detection limit and treated as 0.5, which was between 0 and the lower detection limit of 1. Additionally, some samples from the Rocafuerte pilot study were not included due to high incubation temperatures or low dilutions (n=18) that resulted in inaccurate *E. coli* concentrations. One transect from the Rocafuerte pilot study Day 1 and two transects from Day 2 were ultimately removed from the analysis. The negative controls (n=24) were not included in the analysis. The original, unclean dataset consisted of 397 samples, which resulted in 355 samples after cleaning. For the final analysis, the data was further subsetted to remove samples take downstream (n=13), which resulted in 342 samples for the analysis.

Variables of Interest

New variables were created to simplify the analyses such as the logarithmic value of *E. coli* concentration, velocity (dichotomized as moving (>0 ft./sec) vs not moving (0 ft./sec)), absolute value of incubator temperature difference, and independent unique transects. Using the PROC UNIVARIATE procedure in SAS, the minimum, maximum, median, and mean values were defined for all physicochemical water quality parameters to create a summary table. The geometric mean was used to define the mean values for the *E. coli* concentrations. If the histogram distribution was not normally distributed, median values were used; this was the case for river width, velocity, and turbidity. If the histogram distribution was normally distributed, mean values were used; this was the case for temperature, pH, and DO. The data for pH samples were not used in the analysis due to a faulty probe. Stratification occurred at the river and village level.

Boxplots were created to distinguish the differences in water quality parameters for all three rivers and support the frequency values determined in the summary tables. The graphics were produced using the GRAPH BOX command.

Simple linear regression was conducted to gain insight into the relationships between (1) log *E. coli* concentration and water quality parameters; and, (2) between water quality parameters and distance from shore. Simple linear regression plot information was obtained using the PROC REG procedure on all water quality parameters and distance from shore against log *E. coli* concentration at the river, village, and transect levels. A focus on the river and village analyses further guided the analysis. Log *E. coli* was plotted against turbidity, stream velocity, temperature and DO. Additional scatter plots were created where each water quality parameter was plotted against distance from shore (m) to visualize variability between each river. All plot graphics were produced using the TWOWAY scatter command.

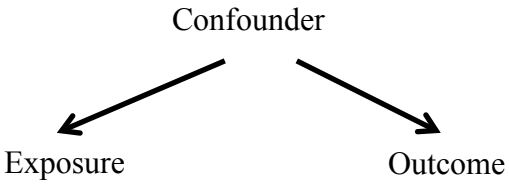
T-tests were conducted on eddy location (coded as 'within' and 'outside') for the Santiago, Cayapas, and Santiago and Cayapas combined. This analysis was executed using the PROC TTEST procedure.

Correlated Mixed Modeling Approach— To expand on the univariate analysis, correlated modeling process followed next. All data was correlated based on the sampling procedure, where multiple samples were taken from the same point over a four-day period, respectively for each village. All samples were associated with a unique transect where multiple point samples were taken along a transect; therefore, the cluster size is the number of unique transects (k=15). The procedure PROC MIXED in SAS was used to

build the model that included log concentration of *E. coli* as the primary outcome, distance from shore as the primary exposure variable, and included temperature, dissolved oxygen, turbidity, velocity (dichotomized) as potential confounders. Interaction variables were created by multiplying the primary exposure variable with all potential confounders. In order to account for the rivers and transects associated with the samples, both variables were included in the CLASS statement. The toeplitz correlation structure was used and a robust estimator was excluded from the modeling. No specialized random effects, intercepts, or slopes were introduced into the model.

The modeling strategy used to build this model assessed for collinearity, interaction, and confounding. First, collinearity was assessed using a collinearity macro. The process of backwards elimination was performed when the condition index (CI) was above 30 and at least two proportions of variance (VDP), not including the intercept, was above 0.5. Next, significant interaction terms were determined using the likelihood ratio test comparing the full model, which included all interaction terms, to the reduced model that did not have any interaction terms. The likelihood ratio test was then used to determine significance at the 95% confidence level for each interaction term. After this process, all confounders and interaction terms remaining were considered the gold standard model. Confounding was then tested by comparing the Odds Ratios (OR) for distance from shore (e^{β}) from the gold standard model to each model that did not include the confounder variable (i.e. reduced model). If the OR for the reduced model was greater than or less than 10% the OR of the gold standard then the variable remained in the model. The final model included all significant confounders and interaction terms at the 95% confidence level.

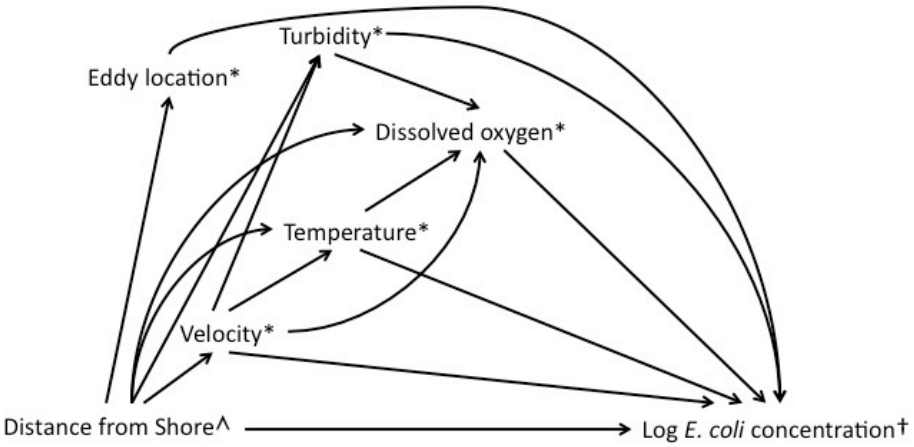
Directed Acyclic Graphs (DAG)— Directed Acyclic Graphs are causal diagrams that explain the relationships among variables in a model. Confounders are causative to both the exposure and the outcome and must be controlled for when building a model. A DAG was created to conceptualize and determine which variables would be considered potential confounders and reduce sources of bias. The basic structure of a DAG is as follows:



DAG for Log E. coli concentration

The Directed Acyclic Graph (DAG) for log *E. coli* concentrations was used to further explain the results of the correlated modeling. The diagram (**Figure 6**) and list of causative pathways confirm that all water quality parameters (velocity, temperature, DO, and turbidity) and eddy location are not potential confounders. All arrows are unidirectional in the model.

Figure 6. DAG on relationships between all exposure, potential confounders and outcome variables of the study. ^Exposure, †Outcome, *Potential confounders



Distance from shore → Temperature → log *E. coli* concentrations

Distance from shore → Velocity → log *E. coli* concentrations

Distance from shore → Turbidity → log *E. coli* concentrations

Distance from shore → DO → log *E. coli* concentrations

Distance from shore → Eddy location → log *E. coli* concentrations

Distance from shore → Temperature → DO → log *E. coli* concentrations

Distance from shore → Velocity → Temperature → log *E. coli* concentrations

Distance from shore → Velocity → Turbidity → log *E. coli* concentrations

Distance from shore → Turbidity → DO → log *E. coli* concentrations

RESULTS

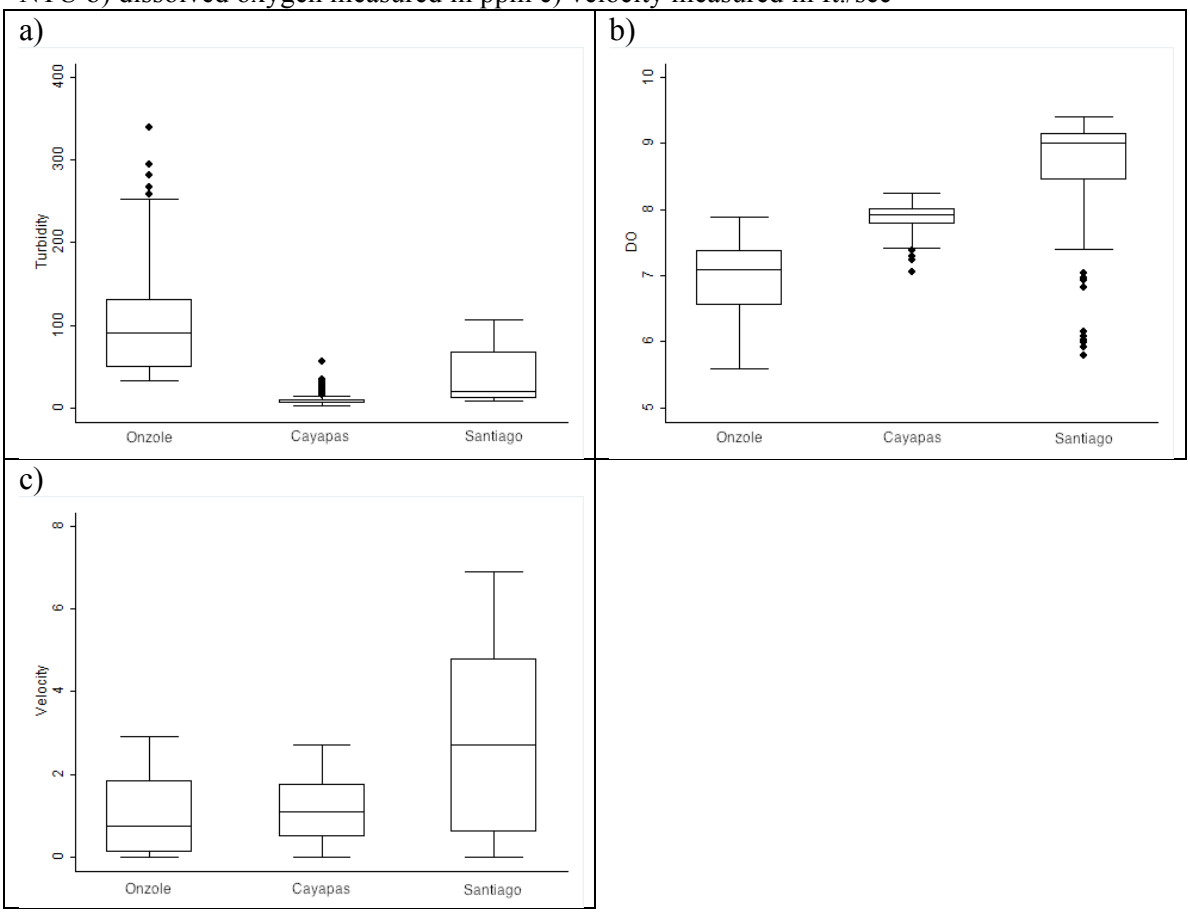
Summary Statistics

Water quality river parameters are summarized by river and village levels in **Table 2** and **Table 3**. Santiago is the fastest flowing river, Cayapas is intermediate, and Onzole is the slowest, and that pattern holds true for all water quality parameters except turbidity. Based on **Table 2**, the highest *E. coli* concentration was for Onzole River with a range of 21.8-15,531 CFU/100mL and a geometric mean of 1247.9 CFU/100mL. The greatest river width was for Santiago River with a range of 18-166m and median value of 144m. The highest temperature was for Onzole River with a range of 24-28.2°C and an average of 26.12°C. The highest velocity was for Santiago River with a range of 0-6.9 ft./sec and a median value of 2.7 ft./sec. The highest dissolved oxygen was for Santiago River with a range of 5.79-9.4 ppm and an average of 8.61 ppm. The highest turbidity was for Onzole River with a range of 32.2-340 NTUs and a median value of 91.05 NTUs. Based on **Figure 2**, the graphics support that Onzole has the highest turbidity and Santiago has the highest DO and velocity measurements. In contrast, Cayapas has the lowest turbidity and Onzole the lowest DO and velocity.

Table 2. Summary Table of Water Quality River parameters by River
Frequency values based on variable distribution as *Geometric mean, +Mean value, and ^Median value.

Rivers	*E. coli** concentration (MPN/100mL)	River width (m)^	Temperature (°C)+	Velocity (ft/sec)^	Dissolved oxygen (ppm)+	Turbidity (NTU)^	Total samples
Onzole	1247.9 (21.8, 5172)	53.5 (35, 66)	26.12 (24, 28.12)	0.6 (0, 2.9)	6.96 (5.6, 7.89)	91.05 (32.2, 340)	120
Cayapas	474.4 (22.3, 4611)	87.5 (73, 93)	25.03 (24.6, 26.1)	1.1 (0, 4.3)	7.87 (7.06, 8.45)	7.03 (3.28, 55.9)	120
Santiago	128 (0.5, 2419.6)	144 (18, 166)	24.85 (23.4, 27)	2.7 (0, 6.9)	8.61 (5.79, 9.4)	19.4 (8.19, 106)	102

Figure 2. Boxplots of water quality parameters stratified by river. a) turbidity measured in NTU b) dissolved oxygen measured in ppm c) velocity measured in ft./sec



Based on **Table 3**, all villages on the same river are similar to one another. The highest *E. coli* concentration was for Tangare, on the Onzole River, with a range of 317-15,531 CFU/100mL and a geometric mean of 1408.3 CFU/100mL. The greatest river width was for La Peña along the Santiago River with a range of 128-166m and median value of 160m. The highest temperature was for Tangare on the Onzole River with a range of 25.4-28.2°C and an average of 26.42°C. The highest velocity was for La Peña on the Santiago River with a range of 0-6.9 ft./sec and a median value of 3.6 ft./sec. The highest dissolved oxygen was also for La Peña with a range of 7.71-9.4 ppm and an

average of 9.05 ppm. The highest turbidity was for Arenales village on the Onzole River with a range of 38.1-340 NTUs and a median value of 101.1 NTUs.

Table 3. Summary of Water Quality River parameters by Village
Frequency values based on variable distribution as *Geometric mean, +Mean value, and ^Median value.

Village	River	*E. coli** concentr-ation (MPN/ 100mL)	River width (m)^	Temp (°C)+	Velocity (ft./sec)^	Dissolved oxygen (ppm)+	Turbidity (NTU)^	Total sample
Arenales	Onzole	1040.9	42	25.69	1.7	7.44	101.1	48
Tangare	Onzole	1408.3	57	26.42	0.2	6.65	87.85	72
Telembi	Cayapas	561.6	81	25.15	1	7.93	6.23	48
Trinidad	Cayapas	424.0	89	24.95	1.1	7.84	7.61	72
Rocafuerte	Santiago	104.6	24	25.64	0.8	7.63	21	30
La Peña	Santiago	139.3	160	24.5	3.6	9.05	15.85	72

Effect of eddy location in Rivers

In order to better understand the effects of eddy location in rivers, all three rivers were stratified separately. For this analysis, the river Onzole was excluded because all eddy locations were 'within' the eddy and did not have an 'outside' eddy location for comparison. Using the Student's t-test and based on the results of **Table 4,** both Santiago and Cayapas did not show a significant difference between observed eddy locations when stratified independently. However, when Santiago & Cayapas were combined, there was a slight statistical significance at the 95% confidence level between eddy locations. These results therefore appear to be driven by sample size, so eddy location was not considered in the correlated modeling and velocity was dichotomized as moving and non-moving water instead.

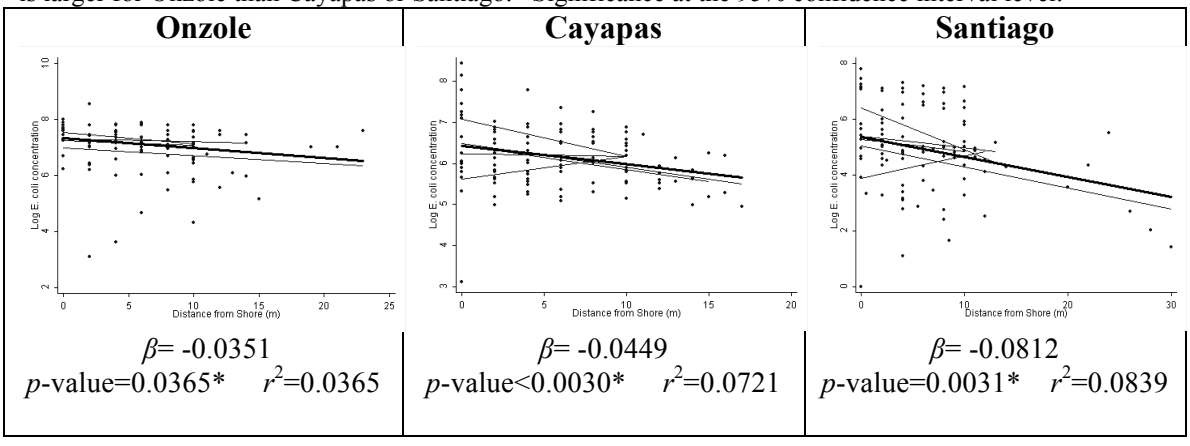
Table 4. Comparison of eddy location for Santiago & Cayapas combined
Data presented represent a significance of eddy location based on river stratification. Student's t-test of the mean values of eddy location was used to test significance. *Indicates marginally significant difference between eddy locations.

River	Eddy location	N	Mean	95% CL Mean	Std. Dev.	p-value
Santiago & Cayapas	Outside	84	5.3557	(5.0747, 5.6368)	1.2951	0.0448*
	Within	137	5.7314	(5.4922, 5.9705)	1.4155	

Comparison of log E. coli concentrations between all three rivers

With different characteristics between rivers, log *E. coli* concentrations were stratified by river. There was a significant association between increased distance from shore and decreased log *E. coli* concentrations for all three rivers in the study (**Figure 3**). The Santiago River demonstrated the strongest association (β= -0.0812, p=0.0031), followed by the Cayapas River (β= -0.0449, p<0.0030), and the Onzole River (β= -0.0351, p=0.0365). However, all three rivers had a poor linear fit (r^2 <0.07).

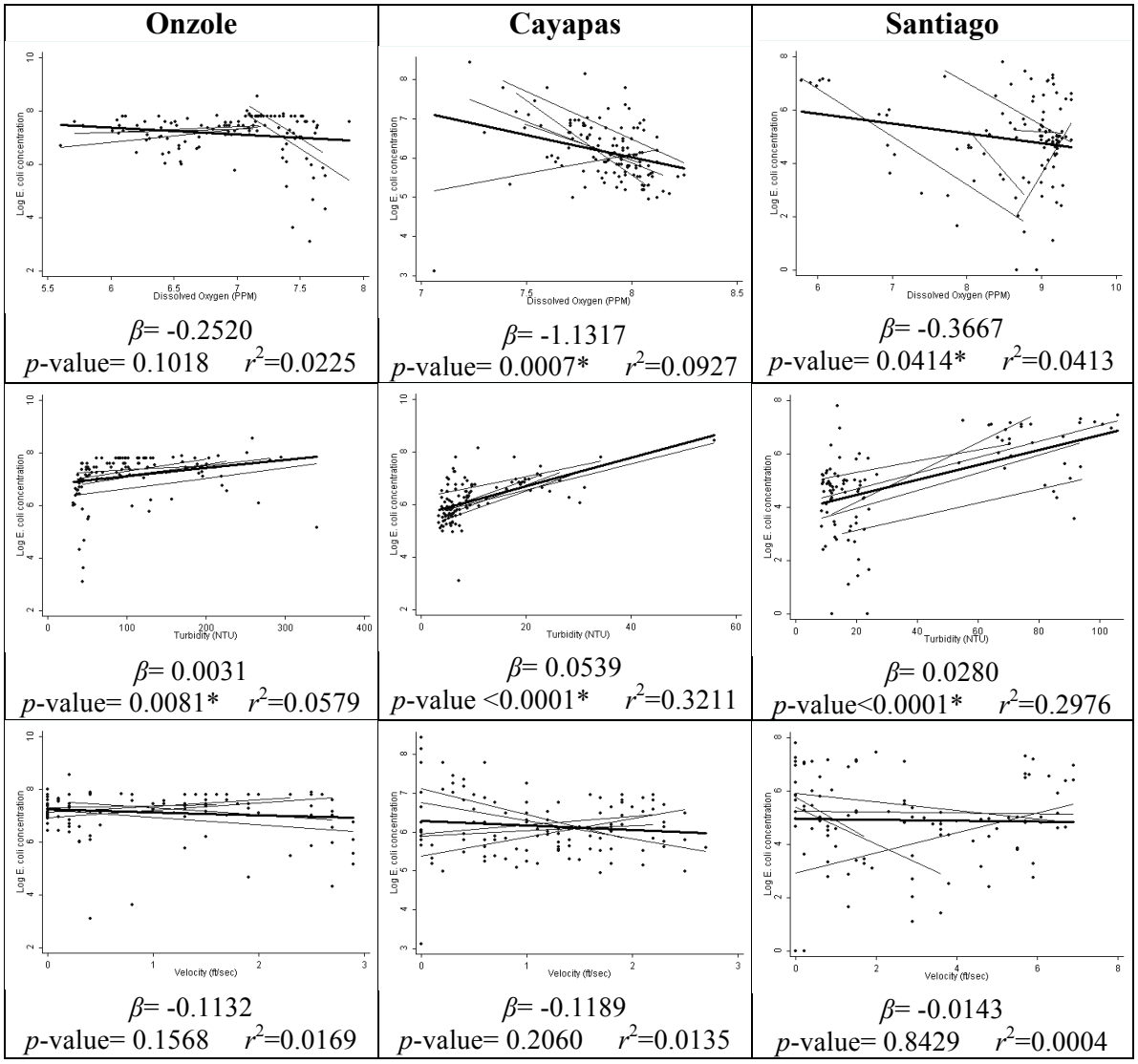
Figure 3. Overall trend of log *E. coli* concentration by River. All three graphs demonstrate negative correlation between log *E. coli* concentration and increased distance from shore. Best-fit lines are noted for each transect and river. The light grey lines are trend lines for individual transects. The solid black lines are trend lines for all transects combined for each respective river. Note: the scale on the y-axis is larger for Onzole than Cayapas or Santiago. *Significance at the 95% confidence interval level.



Comparison of water quality parameters to log E. coli concentrations by River

In order to understand the role of water quality parameters all measured river water parameters were graphed (**Figure 4**). There was a significant association between increased dissolved oxygen and decreased log *E. coli* concentrations for the Cayapas and Santiago Rivers (Cayapas β= -1.1317, p=0.0007; Santiago β= -0.4133, p=0.0282); although, the goodness of fit was not very high for either the Cayapas (r^2=0.0927) or Santiago (r^2=0.0473). The Onzole River did not show any linear relationship between DO and *E. coli* (p=0.1018, r^2 = 0.0225). All three rivers showed a significant association between increased turbidity and log *E. coli* concentrations where Cayapas (β= 0.0539) and Santiago (β= 0.0288) had the strongest associations (both p<0.0001). However, both rivers did not have good linear fit (Cayapas r^2 =0.3211; Santiago r^2 =0.2817). When a simple linear regression was conducted between velocity and log *E. coli* concentrations, there was no significant association for any of the three rivers (all p>0.05).

Figure 4. Overall trend of log *E. coli* concentrations and water quality parameters by River There is a positive relationship between increased E. coli concentrations and increased turbidity, and decreased E. coli concentrations with increased DO. The light grey lines are the trend lines for individual transects. The solid black lines are the trend lines for all transects combined. Note: the scale on the y-axis is larger for Onzole than Cayapas or Santiago.
*Significance at the 95% confidence interval level.

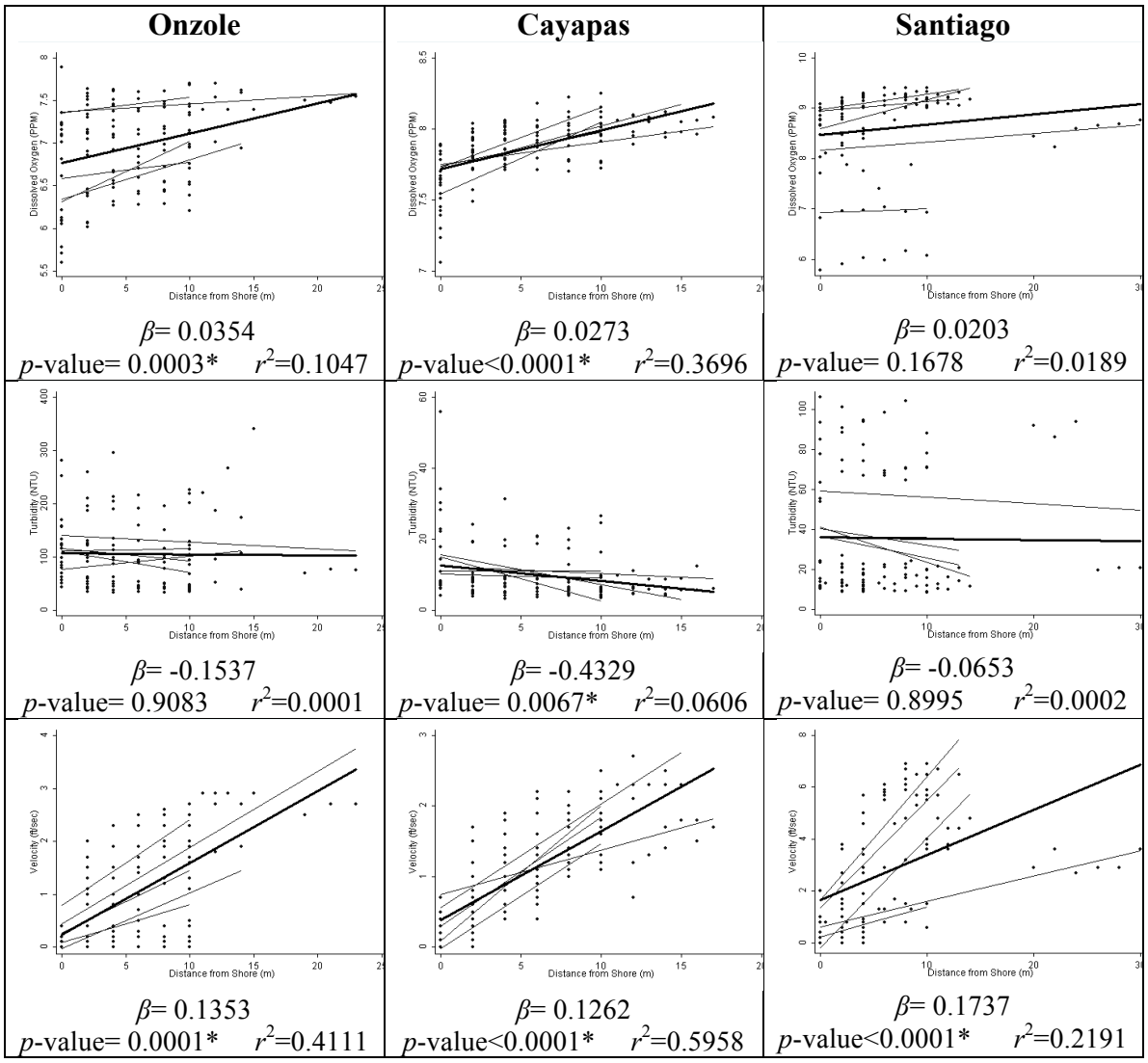


Onzole	Cayapas	Santiago
β= -0.2520	β= -1.1317	β= -0.3667
p-value= 0.1018 r^2=0.0225	p-value= 0.0007* r^2=0.0927	p-value= 0.0414* r^2=0.0413
β= 0.0031	β= 0.0539	β= 0.0280
p-value= 0.0081* r^2=0.0579	p-value <0.0001* r^2=0.3211	p-value<0.0001* r^2=0.2976
β= -0.1132	β= -0.1189	β= -0.0143
p-value= 0.1568 r^2=0.0169	p-value= 0.2060 r^2=0.0135	p-value= 0.8429 r^2=0.0004

Comparison of water quality parameters to distance from shore by River

In order to understand the effect of water quality parameters due to distance from shore scatterplots were created (**Figure 5**). There was a significant association between increased distance from shore and increased DO for the Cayapas and Onzole Rivers in the study (Cayapas β= 0.0273, p<0.0001; Onzole β= 0.0354, p=0.0003). The goodness of fit was not very high for either the Cayapas (r^2=0.3696) or Onzole (r^2=0.1047), although Cayapas showed a better fit for DO. The Santiago River did not show any significant linear relationship between distance from shore and DO (β= 0.0203, p=0.1678, r^2 = 0.0189). Cayapas showed a significant association between increased turbidity and increased distance from shore (β= -0.4329, p=0.0067), where Onzole and Santiago had no associations (both p>0.8995). However, Cayapas did not have good linear fit (Cayapas r^2 =0.0606). Contrary to the results between velocity and log *E. coli* concentrations, there was a strong association between increased distance from shore and increased velocity for all three rivers (all βs>0.1, all $p\leq$0.0001).

Figure 5. Overall trend of water quality variables compared to distance from shore There is a positive relationship between increased distance from shore and increased velocity for all three rivers. The light grey lines are the trend lines for individual transects. The solid black lines are the trend lines for all transects combined. *Significance at the 95% confidence interval level



Onzole	Cayapas	Santiago
β= 0.0354	β= 0.0273	β= 0.0203
p-value= 0.0003* r^2=0.1047	p-value<0.0001* r^2=0.3696	p-value= 0.1678 r^2=0.0189
β= -0.1537	β= -0.4329	β= -0.0653
p-value= 0.9083 r^2=0.0001	p-value= 0.0067* r^2=0.0606	p-value= 0.8995 r^2=0.0002
β= 0.1353	β= 0.1262	β= 0.1737
p-value= 0.0001* r^2=0.4111	p-value<0.0001* r^2=0.5958	p-value<0.0001* r^2=0.2191

Factors associated with log E. coli concentrations in river water samples

The most significantly associated factor with log *E. coli* concentrations is distance from shore, which was considered our primary exposure of interest. The distance from shore*temperature, temperature, and distance from shore*DO variables dropped during the collinearity assessment. The remaining variables were: distance from shore, turbidity, DO, velocity, distance from shore*turbidity, and distance from shore*velocity. The global likelihood ratio test was significant ($p<0.05$) comparing the full and reduced model where Degrees of Freedom=2. Then standard backwards elimination concluded that all interaction terms should be dropped from the model. The gold standard resulted in: distance from shore, turbidity, DO, and velocity (dichotomized) (**Table 5)**. All potential confounders fell out from the model, resulting in the primary exposure variable of distance from shore as the only variable in the model. Model results are in Table 5 below.

Gold Standard equation:
$(\log E.\ coli)_{ij} = \beta_0 + \beta_1 \text{Distance from shore}_{1ij} + \beta_2 \text{Turbidity}_{2ij} + \beta_3 DO_{3ij} + \beta_4 \text{Velocity}_{4ij} + e_{ij}$

Final Model equation:
$(\log E.\ coli)_{ij} = \beta_0 + \beta_1 \text{Distance from shore}_{1ij} + e_{ij}$

Where i= 1-k; k= 15 for number of transects; and j= 1-6 points for each transect.

Table 5: Gold Standard and Final correlated mixed model assessing factors associated with distance from shore. Samples were correlated at the transect level.

	Parameter	Estimate	Std Error	*p*-value
Gold Standard: outcome is log *E. coli*	Intercept	10.5144	1.0267	<0.0001
	Distance from Shore	-0.03023	0.0097	0.0019
	Turbidity	0.006247	0.0014	<0.0001
	DO	-0.5939	0.1352	<0.0001
	Velocity	0.01197	0.1400	0.9319
Final model: outcome is log *E. coli*	Intercept	6.3603	0.2667	<0.0001
	Distance from shore	-0.04883	0.008444	<0.0001

DISCUSSION

Efforts to understand the microscale dynamics of water resources that are potential points of consumption warrant this study, particularly because spatial variability in microbiological water quality has not been extensively explored. This study explores this issue through three different rivers in northern Ecuador.

The primary result was the final global correlated model expressed that there was a significant association between increased distance from shore (exposure; β= -0.04883) and decreased *E. coli* levels (outcome). The final model did not include any significant interaction terms that were interpretable or any significantly measured confounders (Table 5). These results were also supported with the DAG for log *E. coli* concentrations since there were no arrows that were causative associations to distance from shore and log *E. coli* concentrations (Figure 6). This association between increased distance from

shore and decreased *E. coli* held for all three rivers (Figure 3), with the strongest effect in the fastest flowing river (Santiago). Although the model did not show interaction with velocity or turbidity, we did observe a dampened effect in the Onzole versus Cayapas & Santiago (Figure 4). Onzole is higher in turbidity and slower in velocity (Figure 2). There is also a dose-response along the spectrum of DO, turbidity, and velocity of the three rivers (Figure 4). This spectrum is likely due to the differences between the three rivers.

The secondary result from this study is that while the water quality variables were not included in the final model where distance from shore was our primary interest, we observed relationships between all water quality parameters being associated with *E. coli,* which indicates characteristics of these rivers. There are different strengths of effects between the rivers. The strongest association is with turbidity where, at all different turbidity ranges, we observe a clear association between increased turbidity and increased *E. coli* levels (Figure 4). There was less effect than expected (inside mean=5.4 vs. outside mean=5.7) of eddy location on difference in *E. coli* concentration. This could be due to the observational categorization of still water versus flowing water. Additionally, when there was no flow for the Onzole river then all samples were considered within the eddy, which is difficult to use when there is no comparison (i.e. “outside the eddy”).

There were a number of limitations that could have influenced the *E. coli* concentrations in this study. In terms of the modeling, although the final model did not include any measured confounders, but it is still possible that there are other unknown confounders that can and should be included, such as rainfall. One possible explanation for the increased bacterial concentrations closer to shore could be a potential *E. coli*-sediment interaction, where the banks will have shallow sediment and the further out

there will be less sediment and therefore less *E. coli* attached. Previous studies suggest that sediment disturbance can account for most of the total fecal contamination and that re-suspension of microbes from the sediment bed can be a major source for the water column (Nagels, Davies-Colley et al. 2002; Rehmann and Soupir 2009). Since we did not run transects where no human activity occurred we cannot distinguish the sediment attachment process from human contamination. Further investigation of seasonal variations, particularly with rainfall, could influence *E. coli* densities thereby altering the outcome (Lyautey, Lu et al. 2010). Additional storm and seasonal data would allow for a more long-term assessment of stream bacterial quality (Traister and Anisfeld 2006).

CONCLUSION

Our results suggest that increased distance from shore is associated with lower *E. coli* contamination. This result has three major implications. For this primary result: (1) collecting water further from shore may reduce the fecal contamination of surface drinking water sources. Though, it may be more dangerous for children to collect water further out from shore, it is known to have decreased fecal contamination. Further research is needed to understand whether this behavior to collect water further from shore is culturally acceptable; (2) These results also suggest that within a given river, contamination between villages is not critically important because of die-off as distance increases from the shore. The levels of local contamination occur along the transect regardless of village or river; (3) However it is important to keep in perspective that river characteristics are also very important in determining *E. coli* levels. Across all rivers

there was a decrease in *E. coli*, regardless of the different physicochemical differences between the rivers.

This result has further implications, and perhaps future studies, for different water treatment methods such as pre-filter before chlorinating, etc. based on the character of the river. Future investigations should consider *E. coli* enumeration of soil and sediment in rivers of frequent human interaction.

REFERENCES

Begum, Y. A., K. A. Talukder, et al. (2005). "Enterotoxigenic Escherichia coli isolated from surface water in urban and rural areas of Bangladesh." Journal of clinical microbiology **43**(7): 3582-3583.

Byappanahalli, M., M. Fowler, et al. (2003). "Ubiquity and Persistence of Escherichia coli in a Midwestern Coastal Stream." Applied and Environmental Microbiology **69**(8): 4549-4555.

Edberg, S. C., E. W. Rice, et al. (2000). "Escherichia coli: the best biological drinking water indicator for public health protection." Journal of Applied Microbiology **88**: 106S-116S.

EPA (2010). Sampling and Consideration of Variability (Temporal and Spatial) For Monitoring of Recreational Waters. Chapter 3: Findings on Indicator Density and Spatial Variability 23-40.

Fries, J. S., G. W. Characklis, et al. (2008). "Sediment-water exchange of Vibrio sp. and fecal indicator bacteria: implications for persistence and transport in the Neuse River Estuary, North Carolina, USA." Water Research **42**(4-5): 941-950.

Hamner, S., A. Tripathi, et al. (2006). "The role of water use patterns and sewage pollution in incidence of water-borne/enteric diseases along the Ganges river in Varanasi, India." International Journal of Environmental Health Research **16**(2): 113-132.

Heathwaite, A. L. (2010). "Multiple stressors on water availability at global to catchment scales: understanding human impact on nutrient cycles to protect water quality and water availability in the long term." Freshwater Biology **55**(1): 241-257.

Hellweger, F. L. and P. Masopust (2008). "Investigating the Fate and Transport of Escherichia coli in the Charles River, Boston, Using High-Resolution Observation and Modeling " Journal of the American Water Resources Association **44**(2): 509-522.

Ishii, S. and M. J. Sadowsky (2008). "Escherichia coli in the Environment: Implications for Water Quality and Human Health." Microbes and Environment **23**(2): 101-108.

Levy, K., A. E. Hubbard, et al. (2009). "Drivers of Water Quality Variability in Northern Coastal Ecuador." Environmental Science and Technology **43**(6): 1788-1797.

Levy, K., K. Nelson, et al. (2012). "Rethinking Indicators of Microbial Drinking Water Quality for Health Studies in Tropical Developing Countries: Case Study in Northern Coastal Ecuador." American Journal Tropical Medicine and Hygiene **86**(3): 499-507.

Lyautey, E., Z. Lu, et al. (2010). "Distribution and Diversity of Escherichia coli Populations in the South Nation River Drainage Basin, Eastern Ontario, Canada." Applied and Environmental Microbiology **76**(5): 1486-1496.

Nagels, J. W., R. J. Davies-Colley, et al. (2002). "Faecal contamination over flood events in a pastoral agricultural stream in New Zealand. ." Water science and technology: a journal of the International Association on Water Pollution Research **45**(12): 45.

Quilliam, R., K. Clements, et al. (2011). "Spatial variation of waterborne Escherichia coli - implications for routine water quality monitoring." Journal of Water and Health **9**(4): 734-737.

Rehmann, C. R. and M. L. Soupir (2009). "Importance of interactions between the water column and the sediment for microbial concentrations in streams." Water Research **43**(18): 4579–4589.

Strauch, A. M. and A. M. Almedom (2011). "Traditional Water Resource Management and Water Quality in Rural Tanzania." Human Ecology **39**: 93–106.

Traister, E. and S. C. Anisfeld (2006). "Variability of indicator bacteria at different time scales in the Upper Hoosic River watershed." Environmental Science and Technology **40**(16): 4990-4995.

UNICEF and WHO (2012). Progress on Drinking Water and Sanitation: 2012 Update**:** 2-58.

USGS (2012). "How much water is there on, in, and above the Earth?". Retrieved January 8th, 2013, from ga.water.usgs.gov/edu/wusw.html.

WHO (2000). "Water Sanitation Health: Water-related diseases." Retrieved January 8th, 2013, from http://www.who.int/water_sanitation_health/diseases/diarrhoea/en/.

WHO (2009). "Diarrhoeal Diseases." Retrieved January 8th, 2013, from http://www.who.int/mediacentre/factsheets/fs330/en/index.html.

WHO/UN-Water (2008). UN-WATER GLOBAL ANNUAL ASSESSMENT OF SANITATION AND DRINKING-WATER, World Health Organization.

WWAP. "Water Use." UN-Water. from http://www.unwater.org/statistics_use.html.