Article:

Fejfar, D., Tracy, W., Kelly, E.R. et al. (5 more authors) (2024) Identifying predictors of E. coli in rural household water in sub-Saharan Africa using elimination regression. Environmental Science: Water Research & Technology, 10 (5). pp. 1147-1159. ISSN 2053-1400

https://doi.org/10.1039/d3ew00915g

View Article Online
DOI: 10.1039/D3EW00915G

Identifying predictors of *E. coli* in rural household water in sub-Saharan Africa using elimination regression

Donald Fejfar[1,2], Wren Tracy[3], Emma Kelly[1,4], Michelle Moffa[1], Robert Bain[5], Jamie Bartram[1,6], Darcy Anderson[1], Ryan Cronk[1,*]

[1] The Water Institute, Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 135 Dauer Drive, CB #7431, Chapel Hill, North Carolina 27599, United States
[2] Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 3101 McGavran-Greenberg Hall, Chapel Hill, North Carolina 27599, United States
[3] ICF, 2635 Meridian Pkwy Suite 200, Durham, North Carolina, 27713, United States
[4] The Water Project, PO Box 3353, Concord, New Hampshire, 03302, USA
[5] Regional Office for the Middle East and North Africa, UNICEF, Amman, Jordan
[6] School of Civil Engineering, University of Leeds, Leeds, United Kingdom

* Corresponding author. The Water Institute at UNC, Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 135 Dauer Drive, CB#7431, Chapel Hill, NC 27599, United States, *e-mail address:* rcronk@alumni.unc.edu; rcronk@live.unc.edu (Ryan Cronk).

Abstract

Exposure to fecally contaminated drinking water contributes to the global disease burden, especially in sub-Saharan Africa (SSA). We used cross-sectional data and elimination regression analysis to examine factors influencing *E. coli* contamination in household drinking water samples from 4,499 rural households in nine countries in SSA (Malawi, Mozambique, and Zambia in Southern Africa; Ghana, Mali, and Niger in Western Africa; and Kenya, Rwanda, and Tanzania in Eastern Africa). The proportion of household water samples containing *E. coli* was 71%, ranging from 45% (Malawi) to 89% (Tanzania). Pooled and multi-country predictive logistic regression models showed that using an unimproved-type water source, the absence of a community water committee, and domestic animal ownership were significantly associated with household drinking water contamination. Household water treatment and storage practices, sanitation and hygiene practices, and payment for drinking water were not significantly associated with *E. coli* contamination in any model. The season was a significant predictor of *E. coli* in the pooled model; samples collected in the rainy season were 2.3 [2.0, 2.7] times as likely to be contaminated with *E. coli*. Practitioners and policymakers should prioritize implementing piped on-plot water services, establishing effective local water source management structures, and incorporating animal husbandry practices into water, sanitation, and hygiene interventions.

View Article Online
DOI: 10.1039/D3EW00915G

Water Impact Statement

This study demonstrates the importance of local context when planning interventions and assessments related to microbial contamination of household drinking water. Practitioners should prioritize expanding on-plot water services, establishing effective water source management, and incorporating safe animal husbandry practices into water quality interventions. Researchers should investigate the effects of user payment, household water treatment, and storage practices on water quality.

Introduction

Drinking water is an important fecal-oral disease transmission pathway.[1–3] Exposure to fecally contaminated drinking water causes diarrheal diseases and enteric fevers, long-term developmental effects, and death.[4–6] The burden of waterborne disease is most pronounced in low- and middle-income countries (LMICs), including sub-Saharan Africa (SSA), where diarrheal disease is a leading cause of death in children.[7] The United Nations Sustainable Development Goal (SDG) target 6.1 calls for "universal and equitable access to safe and affordable drinking water for all." This target necessitates using "safely managed" drinking water services, which are on-plot, available when needed, and free from fecal and chemical contamination.[8] More than half of the global population lacking safely managed drinking water lives in SSA; an estimated 31% of the population of SSA had access to safely managed drinking water services compared to 73% of the global population in 2022.[9] These challenges are especially pronounced in rural areas of SSA, where 15% of the population has access to safely managed drinking water compared to 53% in urban areas.[10,11]

Escherichia coli (E. coli), a bacterial species found in the gut of humans and other animals, is commonly used to indicate human and non-human fecal contamination, poor sanitation and hygiene conditions, and other disease-causing pathogens in water.[12] Up to half of the population of SSA is exposed to drinking water that does not meet the WHO guidelines for safe drinking water (no detection of *E. coli* in 100 milliliters of drinking water) at a given time. Drinking water is often fecally contaminated in rural and low-resource settings and from unimproved water sources such as surface water, unprotected dug wells, and unprotected springs.[6,13] However, improved water sources (which have the potential to deliver safe water by nature of their design and construction) can also be contaminated by pathways such as open defecation, animal access to water sources, and the deterioration of water quality between source and consumption.[6,14–16] Though it is a commonly reported indicator for tracking progress towards SDG target 6.1, water source type correlates poorly with the absence of microbial contamination.

Understanding the influence of household risk factors on fecal contamination of drinking water can enhance the monitoring of safe water access and inform interventions to reduce human exposure. Numerous household factors beyond water source type reportedly affect household drinking water quality, including household wealth, sanitation and hygiene practices, and the presence of animals and animal feces.[16–19] Broader community and climatic factors – such as rurality, seasonality, and community-level sanitation or waste disposal practices – may have a greater impact on household water quality.[19,20]

View Article Online
DOI: 10.1039/D3EW00915G

While previous studies offer insights into the factors that may affect drinking water contamination, large-scale multi-country studies comparing the factors affecting contamination across contexts are rare. To date, most have been small-scale in a single country,[15,18-19] or rely on large-scale national surveys (e.g., Demographic and Health Surveys [DHS] or Multiple Indicator Cluster Surveys [MICS]) that contain some relevant determinants but are not designed to explore drinking water quality or underlying determinants in depth.[14-15,17] Systematic reviews and other studies with aggregated reporting are limited by variations in definitions and measurement techniques for dependent and independent variables, which makes robust comparison across settings challenging.[12,15]

For this study, we analyzed data from a cross-sectional survey of 4,499 households in nine countries in SSA to examine factors that influence the presence of *E. coli* in rural household drinking water and contextualize findings specific to three regional African Union (AU) groupings. Our findings represent one of the largest multi-country observational datasets designed to examine water, sanitation, and hygiene outcomes (WaSH) and related predictors. Questions were translated into the local language but were identical across contexts, facilitating multi-country comparisons.

Our findings are built on available evidence by analyzing and comparing predictors of water quality across countries. Understanding how different risk factors influence the type and magnitude of fecal contamination across contexts can further inform specific water quality policies, practices, and research in SSA.

Methods

Setting

The Water Institute at the University of North Carolina, Chapel Hill (UNC), and the international non-governmental organization (INGO) World Vision collected data on WaSH in rural areas of 14 LMICs as part of a program evaluation. Here, a subset of data from rural households in nine countries in SSA – all surveyed households with complete water quality data — were analyzed in a secondary analysis of drinking water contamination and associated risk factors. The study included three countries in each included AU region: Malawi, Mozambique, and Zambia in Southern Africa; Ghana, Mali, and Niger in Western Africa; and Kenya, Rwanda, and Tanzania in Eastern Africa. The de-identified datasets from the evaluation are publicly available.[21] For detailed data collection methods, see the program evaluation final report.[22] Methods are summarized below.

Sampling

Cluster-randomized, population-based sampling was used. Between 100 and 200 household subnational administrative sampling units were obtained from the national census bureau or national statistics office of each country. Clusters were randomly selected with probability proportional to size allocation and stratified by World Vision program and comparison areas.

View Article Online
DOI: 10.1039/D3EW00915G

Fifty-six clusters were selected from each stratum. All households were mapped within each cluster, and 25 were randomly selected for the survey.

Up to five households per cluster with drinking water available during the survey were randomly selected for microbial water quality testing. If fewer than five households had drinking water available, all were sampled. All households with microbial water quality testing data and complete survey data were included in this analysis. Clustering and weighting were not accounted for in the study because probability proportional to size allocation produced an equal probability of selection of roughly equally sized clusters and, therefore, households.

Survey design

The survey instrument included interview questions and direct observations of the household environment and characteristics, drinking water (source, storage, and management), sanitation, and hygiene (Table 1). Household surveys were administered to the female head of household, if available.

Local research consultants or World Vision staff translated and verified the survey instrument. Surveys were programmed in mWater (New York, NY, USA), a mobile survey tool. Quality assurance and quality control measures, such as permitted value ranges and automatic GPS recording, prevented enumerators from entering implausible values. Data were collected during a single season in each country from June to August 2017.

Water quality testing

In selected households, respondents were asked to serve water used for drinking. A 100 mL water sample was tested for *E. coli* using the Compartment Bag Test (CBT) (Aquagenx, Chapel Hill, NC, USA). CBTs were processed immediately following surveys by incubating for either 48 hours at ambient temperatures of 25 to 30°C, 24 hours at ambient temperatures above 30°C, or, if the ambient temperature was below 25°C, samples were stored in a cooler with ice until the end of the day in the field and then placed in a 35 to 37°C incubator for 24 hours. Visible color changes after the waiting period indicated the contamination level of *E. coli* in water samples.[23]

Analysis

For cleaning, data from mWater were exported into Stata/SE 14.2 (College Station, TX, USA). Further cleaning and analysis specific to this study were conducted using SAS Version 9.4 (SAS Institute Inc, Cary, NC, USA). In the regression analysis, water contamination was used as a binary outcome variable, defined as the presence or absence of *E. coli* in 100 mL.[6]

Twenty-four binary predictor variables were selected from the larger dataset for testing in predictive models based on theorized relationships, associations, or potential confounding with the outcome variable according to existing qualitative or quantitative evidence and in consultation with subject-matter experts (Table 1).[16,24] The following variables were defined according to the WHO/UNICEF Joint Monitoring Program for Water Supply, Sanitation, and

View Article Online
DOI: 10.1039/D3EW00915G

Hygiene (JMP): unimproved water, no handwashing facility, and water unavailable.[25] Definitions are provided in Table 1.

All selected predictor variables had adequate variation for validity in a model. Seven predictor variables had responses of "Do not know," "Other," or non-responses recorded for one to fifteen households, which were imputed to the reference level and ignored.

Table 1. Predictor variables related to water, sanitation, hygiene, and household demographics used as predictor variables in analyses of E. coli contamination in household drinking water in nine countries in sub-Saharan Africa

Grouping	Variable	Variable definition	Observed / Reported
Outcome	Water contamination	The most probable number (MPN) of ≥1 *E. coli* per 100 mL household drinking water	Observed
Season	Rainy season	Rainy season at the time of sampling	Observed
Primary water source	Unimproved water	No primary water source or classified as unimproved source (e.g., unprotected dug well or spring, surface water)	Reported
	Discontinuous service	The household does not have 24-hour per-day water service at the water source.	Reported
	Off-plot	Primary water source located off-plot	Reported
	Unavailability	Sufficient water unavailable when needed or less than 12 hours per day	Reported
	30+ min to source	The length of time for one round trip to a water source is 30 minutes or greater.	Reported
	Do not pay	The household does not pay for the use of the primary water source	Reported
	No water committee	Water source not managed by a water committee	Reported
Water storage	No treatment	The household does nothing to drinking water to make it safer	Reported
	Storage	The household stores drinking water in the home	Observed
	Animal access	Storage container within reach of animals	Observed
	Unclean container	The container contains dirt, debris, garbage, fecal matter, etc.	Observed
	No container lid	Storage container not entirely covered by a lid	Observed
	Wide container opening	The storage container has a wide opening	Observed
Household characteristics	Owns animals	Ownership of one or more animals by the household	Reported
	No clothesline	Household currently do not have access to clothesline	Reported
	No dish rack	Household currently does not have access to a dish rack	Reported
	No electricity	The household does not have electricity or a power source	Reported
	No rubbish pit	Households do not have access to a rubbish pit for trash disposal	Reported
	Relative education	Primary school highest education completed by either male or female head of household	Reported
	Wealth quintile	Households are separated into quintiles based on documented wealth	Reported
Sanitation	Visible feces	Visible excreta in the yard (15-meter radius of the household)	Observed
	No hygiene facility	No handwashing facility at the household or limited facility without water and/or soap	Observed
	Open defecation	Open defecation practiced by any member of the household	Reported

We created two predictive models using survey logistic regression procedures for the three included AU regions. We also made two pooled models (with all nine study countries), resulting in eight models in total.

View Article Online
DOI: 10.1039/D3EW00915G

To generate the first model for each region and the pooled data, we used logistic regression to predict the odds of unsafe household water contamination against hypothesized predictor variables.[24] To generate the second model, we created reduced models by employing semi-automatic monitored stepwise backward elimination methods. This method entails starting with the full logistic regression model and then using a standard set of guidelines to reduce the model step-by-step to only the predictor variables with a significant effect in the model. At each step of the elimination process, we manually removed the variable with the lowest partial F-statistic. We checked that the model diagnostics (including convergence) were acceptable and that the Akaike information criterion (AIC) improved.[26,27] We continued manual elimination until all remaining predictor variables had low enough partial F-statistics to produce p-values that were less than the significance-to-stay value ($p \leq 0.05$).

The SAS procedures automatically assessed collinearity. Collinearity was not apparent except for the season variable (whether it was rainy season at the time of sampling). All countries in the Western AU region were surveyed in the rainy season and all other countries in the dry season, resulting in exact collinearity between country and season and between the AU region and season. Because season is an important factor in water quality, we included this variable in the pooled model but did not include it in the regional models.

We used descriptive statistics to determine whether predictor variable odds ratios showed a different relationship with the outcome in the respective country grouping than their observed relationship in descriptive statistics. No instances were found; assessments of potential interactions were insignificant and did not alter effects. We checked residual plots to ensure the link function was appropriate and observed constant variances. We checked ANOVA tests of model fit, convergence criteria, and other automatic model diagnostics plots produced by SAS.

Ethics

This study comprises a secondary analysis of publicly available, de-identified data. When data were collected initially, all participants gave informed consent in their language before enrollment. Data collection was approved by the UNC-Chapel Hill Institutional Review Board (IRB #17-0663) and appropriate agencies within each included country.

Results

The analysis included 4,499 household surveys and drinking water quality results from rural areas in the nine study countries: 1,550 responses in the Southern AU region, 1,369 responses in the Western AU region, and 1,580 responses in the Eastern AU region.

Countries from West Africa were sampled during the rainy season, accounting for 30.4% of all study households (Table 2). Overall, 31.1% (± 0.7%) of households reported using an unimproved primary drinking water source, and 35.5% (± 0.6%) reported open defecation. Most households had no hygiene facilities (91.8% ± 0.4%), stored drinking water (85.2% ± 0.5%), had an off-plot water source (85.0% ± 0.5%), and owned animals (82.3% ± 0.6%) (Table 2). A minority of households did not have drinking water available when needed (6.6% ± 4.4%), had discontinuous water service (18.0% ± 0.6%), reported traveling 30+ minutes to collect water (26.1% ± 0.6%), did not have a lid on water storage containers (28.4% ± 0.7%), or had an unclean water storage container (28.4% ± 0.7%) (Table 2).

Table 2. E. coli in drinking water collected from households (N=4,499) in a rural water program evaluation survey in nine sub-Saharan African countries.

Grouping	Predictor variable	Percent of all households % (standard error)	Percent of *E. coli* in drinking water % (standard error)
Season	Rainy season	30.4 (N/A)	80.4 (1.1)
Main water source	Unimproved water	31.1 (0.7)	80.1 (1.1)
	Discontinuous service	18.0 (0.6)	68.8 (1.6)
	Off-plot	85.0 (0.5)	71.3 (0.7)
	Unavailability	6.6 (0.4)	68.6 (2.7)
	30+ min to source	26.1 (0.6)	72.3 (1.3)
	Do not pay	66.0 (0.7)	72.3 (0.8)
	No water committee	41.0 (0.7)	75.5 (1.0)
Water storage	Storage	85.2 (0.5)	70.7 (0.7)
	No treatment	68.2 (0.7)	70.2 (0.8)
	Animal access	39.1 (0.7)	70.1 (1.1)
	Unclean container	28.4 (0.7)	71.4 (1.2)
	No container lid	28.4 (0.7)	69.1 (1.3)
	Wide container opening	57.4 (0.7)	70.9 (0.9)
Household characteristics	Owns animals	82.3 (0.6)	72.1 (0.7)
	No clothesline	31.8 (0.7)	74.5 (1.1)
	No dish rack	64.8 (0.7)	72.0 (0.8)
	No electricity	64.1 (0.7)	70.2 (0.8)
	No rubbish pit	52.8 (0.7)	71.7 (0.9)
	Lower relative education	48.3 (0.7)	71.3 (1.0)
	Wealth quintile		
	Poorest	21.0 (0.6)	73.5 (1.5)
	Poor	21.3 (0.6)	72.7 (1.5)
	Middle	20.2 (0.6)	73.4 (1.4)
	Wealthy	19.0 (0.6)	70.1 (1.5)
	Wealthiest	18.4 (0.6)	63.7 (1.5)
Sanitation	Visible feces	53.1 (0.7)	74.4 (0.9)
	No hygiene facility	91.8 (0.4)	70.7 (0.7)
	Open defecation	35.5 (0.6)	75.6 (1.1)
Total			**70.5 (0.6)**

Presence of *E. coli*

E. coli was detected in 70.5% (± 0.6% standard deviation) of all 4,499 100 mL household drinking water samples, ranging from 45% (Malawi) to 89% (Tanzania) of households. Relative to other predictor variables, the proportion of samples with *E. coli* present was highest where sample collection took place during the rainy season (80.4% ± 1.1%, only Western AU) or in households with an unimproved water source (80.1% ± 1.1%). The proportion of samples with *E. coli* present was lowest in households in the wealthiest quintile (63.7% ± 1.5%).

Predictive regression models

All elimination models had higher measured predictive power according to the AIC than unadjusted models (Table 3). No observable trends in residual plots were apparent for all elimination models, ANOVA tests were significant (p<0.0001), and convergence criteria were

View Article Online
DOI: 10.1039/D3EW00915G

met; no other automatic model diagnostic plots indicated concerns with model fit or assumptions. Fifteen variables were significant predictors of the presence of *E. coli* in at least one of the four reduced models (all countries, South AU region, West AU region, and East AU region) (Table 3). Nine predictor variables were not significant predictors in any reduced models: discontinuous service, unavailability, no treatment, storage, animal access to stored drinking water, unclean container, no container lid, no hygiene facility, and open defecation.

For the pooled model that included all nine study countries, seven predictor variables remained significant in the reduced model, and all were associated with increases in the odds of *E. coli* presence in household drinking water. Season (sampling during the rainy season) had the largest effect size of all significant predictor variables in the pooled model; sampling during the rainy season had greater odds of detecting *E. coli* in drinking water samples by 2.3 [2.0, 2.7] times relative to sampling during the dry season.

Using an unimproved-type water source was the next largest effect size in the pooled model; the odds of detecting *E. coli* were 1.9 [1.6, 2.3] times higher in households with an unimproved-type water source compared to households with improved-type water sources. The remaining variables significant in the pooled model were owning animals (1.4 [1.1, 1.6]), not having a community water committee (1.3 [1.1, 1.5]), having visible feces in the yard (1.3 [1.1, 1.4]), having a main water source off-plot (1.2 [1.1, 1.5) or 30+ minutes away (1.2 [1.1, 1.4]), or being in the lowest wealth quintile (1.1 [1.1, 1.1]).

Table 3. Odds ratios and confidence intervals for predictor variables for E. coli in household drinking water from unadjusted models and reduced models from backward elimination in nine countries in sub-Saharan Africa.

Category name	Predictor variable	Pooled	Reduced	South Region	Reduced	West Region	Reduced	East Region	Reduced
		OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Season	Rainy season	2.3 (1.9, 2.7)	2.3 (2.0, 2.7)						
Main water source	Unimproved water	1.9 (1.6, 2.3)	1.9 (1.6, 2.3)	1.7 (1.3, 2.4)	1.8 (1.3, 2.5)	0.9 (0.6, 1.3)		2.0 (1.5, 2.6)	1.9 (1.5, 2.6)
	Discontinuous service	1.0 (0.8, 1.2)		1.2 (0.8, 1.5)		1.3 (0.8, 2.0)		0.7 (0.5, 1.0)	
	Off-plot	1.2 (1.0, 1.5)	1.2 (1.1, 1.5)	1.6 (1.1, 2.2)	1.6 (1.2, 2.3)	1.0 (0.7, 1.5)		1.3 (0.9, 1.9)	
	Unavailability	1.1 (0.8, 1.6)		1.2 (0.7, 2.0)		0.5 (0.2, 1.2)		1.4 (0.9, 2.2)	
	30+ min to source	1.2 (1.0, 1.4)	1.2 (1.1, 1.4)	1.3 (1.0, 1.7)	1.4 (1.1, 1.7)	0.9 (0.6, 1.4)		0.9 (0.7, 1.1)	
	Do not pay	1.0 (0.9, 1.2)		0.7 (0.6, 1.0)	0.7 (0.6, 0.9)	1.3 (0.9, 1.7)		1.5 (1.1, 2.0)	1.5 (1.1, 1.9)
	No water committee	1.2 (1.1, 1.4)	1.3 (1.1, 1.5)	1.6 (1.2, 2.2)	1.7 (1.2, 2.2)	1.6 (1.1, 2.2)	1.6 (1.1, 2.1)	0.8 (0.6, 1.0)	
Water storage	Storage	1.3 (1.0, 1.7)		1.1 (0.7, 1.9)		1.3 (0.7, 2.5)		1.1 (0.8, 1.6)	
	No treatment	0.8 (0.7, 1.0)		0.8 (0.5, 1.1)		1.0 (0.7, 1.5)		1.2 (0.9, 1.7)	
	Animal access	1.0 (0.9, 1.2)		0.9 (0.7, 1.2)		0.9 (0.6, 1.2)		1.1 (0.8, 1.4)	
	Unclean container	1.0 (0.9, 1.2)		1.0 (0.8, 1.3)		1.0 (0.7, 1.5)		0.9 (0.6, 1.2)	
	No container lid	0.9 (0.8, 1.0)		1.0 (0.8, 1.3)		1.1 (0.8, 1.6)		0.9 (0.7, 1.3)	
	Wide container opening	1.0 (0.8, 1.1)		1.0 (0.7, 1.3)		0.9 (0.6, 1.2)		1.6 (1.2, 2.2)	1.8 (1.4, 2.4)
Household characteristics	Owns animals	1.4 (1.2, 1.6)	1.4 (1.1, 1.6)	1.4 (1.1, 1.8)	1.4 (1.1, 1.8)	1.5 (1.0, 2.3)		1.3 (1.0, 1.8)	1.4 (1.1, 1.9)
	No clothesline	1.2 (1.0, 1.3)		1.2 (1.0, 1.6)		1.4 (1.0, 2.0)	1.4 (1.1, 2.0)	0.9 (0.7, 1.2)	
	No dish rack	1.0 (0.9, 1.2)		0.8 (0.7, 1.1)		0.9 (0.6, 1.2)		1.3 (1.0, 1.7)	1.3 (1.1, 1.7)
	No electricity	1.0 (0.9, 1.2)		1.0 (0.8, 1.4)		1.2 (0.9, 1.7)		1.4 (1.1, 1.9)	1.4 (1.1, 1.8)
	No rubbish pit	1.0 (0.9, 1.2)		1.3 (1.1, 1.7)	1.4 (1.1, 1.7)	0.6 (0.5, 0.9)	0.6 (0.5, 0.8)	1.1 (0.8, 1.4)	
	Relative education	1.1 (0.9, 1.2)		1.3 (1.1, 1.7)	1.4 (1.1, 1.7)	0.8 (0.6, 1.1)		0.9 (0.7, 1.1)	
	Wealth quintile	1.1 (1.0, 1.1)	1.1 (1.1, 1.1)	1.1 (1.0, 1.2)		1.1 (1.0, 1.2)	1.2 (1.1, 1.3)	1.0 (0.9, 1.1)	
Sanitation	Visible feces	1.3 (1.1, 1.4)	1.3 (1.1, 1.4)	1.2 (0.9, 1.5)		1.4 (1.0, 1.8)	1.4 (1.1, 1.8)	1.2 (0.9, 1.6)	
	No hygiene facility	1.0 (0.8, 1.3)		1.0 (0.7, 1.6)		1.3 (0.8, 2.0)		0.8 (0.5, 1.3)	
	Open defecation	1.0 (0.9, 1.2)		1.0 (0.8, 1.3)		1.3 (0.9, 1.7)		1.0 (0.7, 1.5)	
Full model characteristics	# households	4499	4499	1550	1550	1369	1369	1580	1580
	AIC	5215	5195	2039	2022	1679	1664	1349	1328
	ANOVA test P-value	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

View Article Online
DOI: 10.1039/D3EW00915G

In the Southern AU region reduced model (Malawi, Mozambique, Zambia), the predictor variables with the largest effect sizes were using an unimproved water source (1.8 [1.3, 2.5]), no water committee (1.7 [1.2, 2.2]), and a water source off-plot (1.6 [1.2, 2.3]). Having a lower education level [1.1, 1.8], no rubbish pit [1.1, 1.7], owning animals [1.1, 1.8], and having a water source 30+ minutes away [1.1, 1.7] were each associated with 1.4 times greater odds of households having *E. coli* in drinking water compared to reference levels. Households that did not pay for water were less likely to have *E. coli* in their drinking water (0.7 [0.6, 0.9]).

In the Western AU region reduced model (Ghana, Mali, Niger), the odds of contamination of household drinking water with *E. coli* were greater by 1.6 [1.1, 2.1] times for households with no water committee; 1.4 [1.1, 1.8] times for households with visible feces; 1.4 [1.1, 2.0] times for households with no clothesline; and 1.2 [1.1, 1.3] times for households in a lower wealth quintile. Contamination of household drinking water was less likely in households with no rubbish pit (0.6 [0.5, 0.8]) than samples from households with a rubbish pit.

In the Eastern AU region reduced model (Kenya, Rwanda, Tanzania), households using an unimproved water source had 1.9 [1.5, 2.6] times greater odds of *E. coli* contamination of drinking water contamination than households with improved water sources. A water storage container with an opening wide enough for people's whole hands to touch the water was associated with 1.8 [1.4, 2.4] times the odds of *E. coli* contamination compared to storage containers with narrow openings that protect the water from hands and objects. The odds of contamination of drinking water in this region also were greater by 1.5 [1.1, 1.9] times for households that did not pay for water, 1.4 [1.1, 1.9] times for households that owned animals, 1.4 [1.1, 1.8] times for households that did not have electricity; and 1.3 [1.1, 1.7] times for households that did not have a dish rack.

Model comparison

In each region, all predictor variables had similar odds ratios in the reduced and unadjusted models. Fifteen predictor variables were significant in at least one reduced model – the odds ratios and associated 95% CIs are shown in Figure 1. Using an unimproved water source, owning animals, and not having a community water committee were significant in the pooled model and two of the three regional models. Having an off-plot water source, requiring 30+ min to collect water, being in the lowest wealth quintile, and having visible feces in the yard were all significant in the pooled and one regional model. Five predictor variables were significant in only one model: lower relative education level was significant in the Southern AU region; having a storage container with a wide opening, no dish rack, and no electricity were all significant in the Eastern AU region, and having no clothesline was significant in the Western AU region.

Two predictor variables demonstrated significant but conflicting associations with *E. coli* in household drinking water across regions. Not paying for water was associated with lower odds of contamination in the Southern AU region (0.7 [0.6, 0.9]), but greater odds of contamination in the Eastern AU region (1.5 [1.1, 1.9]) reduced models. Likewise, not having a rubbish pit was associated with lower odds of contamination in the Western AU region (0.6 [0.5, 0.8]) and

View Article Online
DOI: 10.1039/D3EW00915G

greater odds in the Southern AU region (1.3 [1.1, 1.7]) reduced models. Two insignificant predictor variables in unadjusted models became significant in the elimination process: wealth quintile in the Western AU region and time to source (30+ min away) in the pooled elimination model.

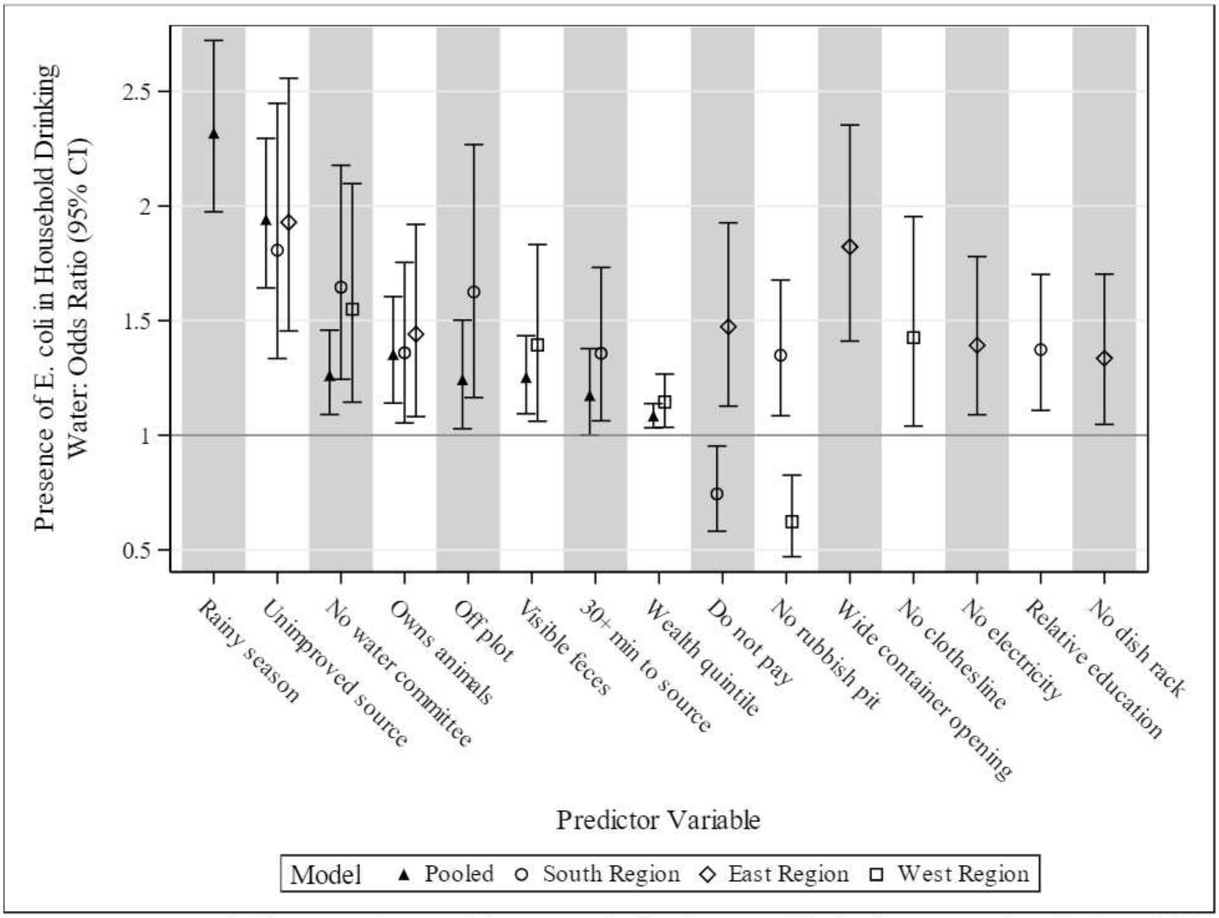

Figure 1. Comparison of odds ratios and 95% confidence intervals of predictor variables for the presence of E. coli in household drinking water from pooled and African Union region logistic regression elimination models.

Discussion

Using cross-sectional survey and drinking water contamination data from 4,499 rural households, we identified factors influencing *E. coli* contamination of household drinking water in nine countries in SSA. The households included in this study had similar but slightly lower WaSH access relative to WHO/UNICEF JMP 2022 estimates for rural households in SSA.[11] Our data are consistent with other studies showing lower access rates in rural versus than urban settings.[16,28] Our results confirm that most rural study households drink water contaminated with *E. coli*.

Our study builds on available evidence by providing insight into how predictors of drinking water quality vary across contexts. A strength of our research is that the same indicator definitions, survey questions, and data collection methods were applied across each country,

View Article Online
DOI: 10.1039/D3EW00915G

facilitating direct comparison across regions with more reliability than in previous studies. Previous studies have examined water quality using similarly large datasets (e.g., MICS) but lack the wide range of predictors, such as water source management, we have discussed here.[19] Understanding the factors that influence fecal contamination and how they change across contexts can inform efforts to increase safe drinking water access, overcome drinking water-related disparities in LMICs, and help improve their monitoring and outcomes. Our results, however, should not be over interpreted – this study was conducted in rural areas of LMICs in SSA and results may vary substantially as compared to urban areas or humanitarian contexts.

Main water source

Most main water source type characteristics significantly influenced drinking water contamination in at least one reduced model. In three of four reduced models – including the pooled model – households with an unimproved-type water source were significantly more likely to have contaminated drinking water than those with an improved-type source. This result is consistent with studies demonstrating improved sources have less *E. coli* contamination and better overall water quality.[11,17,29]

Regional differences in the predictive power of the improved/unimproved variable may be driven in part by the distribution of source types within each classification. For example, in the Western AU region—the only region where improved source type was not predictive of contamination—most unimproved sources were unprotected dug wells (75.9%), with relatively little surface water use (17.9%). In contrast, surface water was the predominant source type in the Eastern and Southern regions among those using unimproved sources (49.7% and 42.6%, respectively).

The improved/unimproved-type classification is a weak predictor of safe water, especially at a single time point. Multiple studies have found that improved-type water sources contain microbial contamination.[30–32] We found that 69% of all households in this study had access to an improved-type water source according to the JMP classification, but *E. coli* was detected in two-thirds (67%) of those households. We did not specifically examine the entry point for contamination (i.e., where contamination was introduced along the chain of collection, transport, and/or storage). However, our results reaffirm that water service level as a stand-alone variable is an imperfect and incomplete proxy for water quality, mainly where households collect drinking water from off-plot sources and then transport and store it in the home.[33,34]

Households accessing water sources that were off-plot or 30+ minutes round trip from the household had a greater likelihood of *E. coli* in the Southern AU region and pooled models. An adverse impact of longer distance or time to source on household water quantity and quality has been previously documented, potentially due to contamination during transportation. However, the relationship is nonlinear and context dependent.[24,28,33] In our study, we observed regional differences in the self-reported average time to collect water (mean round trip collection time in minutes: Western 12.8; Eastern 23.1 minutes; Southern 21.2), which may, in part, drive regional differences in the predictive power of this variable. However, recent reviews and intervention studies in rural Africa have demonstrated that household drinking water quantity did not improve unless the water source was closer to the household and on-plot.[35,36]

We observed greater odds of contamination in households collecting water from a source with no water committee in the pooled Southern AU region and Western AU region models. This is consistent with evidence from other studies conducted in SSA concerning the effects of operational, management, and repair capacities of community water committees.[33,34,37] Community management is a common approach to water system management in rural SSA. Water committees may protect the source from contamination by facilitating system maintenance and repair, managing the environment around the water source (removing foliage, sweeping and keeping out animals), and enforcing rules about source use.[33,34,37] Water committees sometimes educate the community about hygiene and sanitation, which may reduce contamination within the home.[34,37,38] Even as community management of rural water supplies is increasingly critiqued for sustainability reasons,[39,40] water committees can create a sense of community ownership over the water system and drive protective factors.[41] The dataset did not contain measures of the specific activities of water management committees, but the specific structure and activities of committees may drive regional differences.

Paying for water service was significantly associated with increased *E. coli* contamination in the Eastern AU regional model but decreased contamination in the Southern AU region model. Paying for water service is thought to improve water quality by providing funds for repairs that protect the source.[42] However, studies from SSA suggest that in some communities, non-monetary contributions - such as contributing labor or leveraging community assets – may be sufficient for repairs and prevent the need for household financial payment.[38] The differing results between the Southern and Eastern AU models further illustrate the complex relationships between water payment and quality and demonstrate the importance of local context in identifying determinants of household water contamination.

Having water available at the time of the survey and continuous service for 24 hours per day are essential aspects of a high-quality drinking water source.[11,28] However, neither was significant in any of the reduced models. This finding may be explained by the study context, where most households had continuous (82%) and available (93%) drinking water access, and few households relied on intermittent supplies.

Water storage and household water treatment

Contamination of drinking water between source and consumption often occurs due to unsafe household storage practices and a lack of proper household water treatment.[16,43–48] In our models, most water storage and treatment characteristics were not significantly associated with *E. coli* contamination. Water storage, household water treatment, storing water within reach of animals, having an unclean storage container, and not having lids for storage containers were all insignificant in every reduced model. In one model (Eastern AU region), there were increased odds of contamination in households using storage containers with openings wide enough to fit a hand. Trials of safe storage interventions in West Africa have

DOI: 10.1039/D3EW00915G

struggled to achieve high, sustained uptake, and improper storage was commonly observed (e.g., using a communal cup to scoop water instead of dispensing it via the tap).[49,50]

The potential for water contamination after collection varies between settings. Contamination potential is exceptionally high in locations where water is collected off-premises, transported, and stored, as each step offers a new potential entry point for contamination.[14,48,51] The households studied primarily obtained water from unimproved-type water sources, and most households sampled were positive for *E. coli.* Poor source water quality may be masking the relationships between water quality and household storage and treatment in this study, as contamination during household storage and in the absence of treatment are more accessible to identify if there is less contamination at the source. Some household water treatment may effectively decrease *E. coli* contamination but not to below 1 MPN per 100 mL, which was the outcome used in our binary analysis.[52]

Sanitation and hygiene

Evidence linking exposure to fecal matter, drinking water microbial contamination, and diarrheal disease is common.[3,53] Drinking water samples tested positive for *E. coli* more often when enumerators observed visible feces around the household in the pooled and Western AU models. However, practicing open defecation was not a significant predictor in any model. We hypothesize that visible feces in the yard may be a stronger predictor because there is a greater opportunity for exposure when feces are closer to the home than open defecation sites farther away. However, this finding may be biased by the reliability of self-reported open defecation practice compared to direct observation by enumerators. Having no handwashing facility with soap and water was not a significant predictor of fecal contamination in the surveyed households, likely because very few households had such a facility (8% or less in any given country).

Household characteristics

Socioeconomic and demographic characteristics of households are well-established predictors of household water quality; households with low education levels in rural, low-income settings are more prone to water contamination compared to wealthier, more well-educated, urban households.[11,24,28] Studies have demonstrated that poor households are less likely to use improved primary water sources or demand high-quality service and quick repairs.[24,54–56] This analysis accounted for the predictive value of the relative education of the heads of household and the wealth quintile of households relative to all those sampled within each country. In the Southern AU region model, households where the head of household did not complete at least a primary education had significantly higher odds of contamination relative to other houses. In the pooled and Western AU models, households in a lower relative wealth quintile experienced slightly higher contamination odds. Overall, however, the influence of socioeconomic variables was relatively low across the models, likely because all surveyed households were in rural, low-income, and low-resource communities.

View Article Online
DOI: 10.1039/D3EW00915G

Visible human or non-human feces and animal ownership significantly contributed to the ability to predict water contamination in all four reduced models. Household ownership of animals, including pets and livestock, is common in the study region (82.3%). Regardless of on-site sanitation and hygiene practices, non-human feces may contaminate stored water or water consumption-related household items such as cups and basins and present long-term and acute health effects. Previous literature demonstrates that animal ownership contributes to greater fecal contamination of household water in LMICs.[18,57,58] We found that households that owned at least one animal had 1.4 times the odds of drinking water contamination than households without animals (pooled: [1.1, 1.6], Southern AU region: [1.1, 1.8], Eastern AU region: [1.1, 1.9]).

While animal ownership was common in the Western AU region, it was not significantly predictive of contamination. We assessed ownership as a binary variable, but a more nuanced analysis of animal husbandry practices and how they vary by region may elucidate this relationship. While not investigated in this study, evidence suggests that animal fecal contamination risk may be influenced by the type of livestock and context-specific husbandry practices, especially around animal waste management and whether animals enter domestic environments.[18,57,58]

We found that various environmental hygiene measures (e.g., having a dish, clothesline, or rubbish pit) were important across all regions. However, the strength and significance of specific predictors varied by region. This supports other studies that suggest that while fecal-oral disease is common, the exposure profile and associated risk factors vary substantially across contexts.[59] Household access to electricity, a dish rack, a clothesline, or a rubbish pit may be related to overall household resources, like wealth and general household cleanliness. However, little evidence exists on the relationship between these factors and water contamination. Logistic regression models by McGarvey et al. (2008) found that household trash disposal and household access to electricity showed mixed or no effects on water contamination, depending on the model.[24] Kirby et al. (2016) found that community-level open solid waste disposal increases fecal contamination of household water in Rwanda, but this may not be related to rubbish pit ownership at a household level.[20]

Season

Season is a well-established determinant of microbial drinking water contamination and associated diarrheal incidence. Case studies from countries within the region show rural, unimproved sources are especially vulnerable to contamination during the rainy season due to increased runoff, flooding, and limitations on external support or resource mobilization.[16,28,34,60–62] However, one case study from rural South Africa found that even as source water contamination may increase during the wet season, household-level contamination patterns may differ due to confounding factors such as increased water quantity for hygiene and cleaning and decreased water storage times during the wet season.[63]

In our study, only countries in Western Africa were sampled during the rainy season; thus, the season was only included in the pooled regression model. In that analysis, season was the

View Article Online
DOI: 10.1039/D3EW00915G

largest predictor of drinking water contamination in any of the reduced models, with an odds ratio of 2.3 [2.0, 2.7]. This may partially explain why the proportion of households in Western Africa, whose participating countries were sampled in the rainy season, had higher proportions of *E. coli* presence (80%) in drinking water samples than in Eastern (76%) or Southern Africa (56%). However, UNICEF and WHO household survey data also suggest household water quality in Western Africa may be lower than in other regions.[64] Due to the cross-sectional nature of this study, it is not feasible to fully disentangle seasonality from other context-specific factors that resulted in lower water quality in Western Africa.

These results highlight the importance of sampling during the rainy and dry seasons (at a minimum, and ideally more intensively therein) to use the season as a covariate or confounder in other drinking water characteristics analyses. Sampling in both seasons will facilitate exploration of the effect of households using secondary water sources during different seasons.

Programming and policy implications

We recommend that policymakers and practitioners prioritize expanding on-plot water services where feasible, establishing effective local water source management structures, and incorporating animal-rearing practices into WaSH interventions.

Our results agree with previous studies that show that an improved, on-plot water source is the best way to ensure clean water in rural households.[29,65] Safe water storage from off-plot sources is often promoted as an interim solution. Still, our study suggests this may not adequately reduce exposure and frequently lead to contaminated water consumption. Along with improving quality, an on-plot source improves health by increasing household water quantity.[66,67] This change in water quantity is as essential to avoid water-washed diseases as water quality is to avoid waterborne diseases.[67] Expanding on-plot access has often been dismissed in rural areas due to the higher cost of this level of service. Still, economic analysis shows that the increase in household productive time more than offsets the cost of water infrastructure expansion.[68] We recommend policymakers and practitioners prioritize expanding on-plot services over expanding communal sources or improving household water storage and treatment practices, in alignment with the reorientation of policy recommendations from the Millennium Development Goals to the Sustainable Development Goals.

Our results in the Southern and Western AU regions and the pooled model show that the absence of a water committee is associated with water contamination at the household level. In parallel with expanding on-plot services, policymakers and practitioners should support local water source management structures in rural areas.

Finally, these results emphasize the importance of considering animals and livestock in water, sanitation, and hygiene interventions. In this dataset, the presence of domestic animals was a better predictor of *E. coli* than open defecation, confirming previous findings that animal feces may have an equal or greater contribution to fecal contamination of drinking water relative to human feces.[18,69]

View Article Online
DOI: 10.1039/D3EW00915G

There may be simple ways to reduce animal-related contamination of household drinking water: one study demonstrated higher rates of fecal contamination when animals were free-roaming, implying that compound fences or chicken coops could reduce contamination; another found that households were less likely to dispose of animal waste than human waste adequately, and recommended providing feces scoops to encourage more effective compound cleaning.[18,58,70]Animal-rearing practices are a vital area of collaboration between the water, sanitation, and hygiene and food/farming sectors, representing an opportunity to align the interventions of all sectors to a One Health approach.[71]

Research implications

Extensive literature examines the relationship between water contamination and water source characteristics, such as water source type, the accessibility of the water source (time/distance), and the presence of a water committee. However, the effect of household water treatment and storage practices, household characteristics (owning a dish rack, rubbish pit, etc.), and animal ownership on household drinking water quality are inadequately understood. While this study confirms that water source characteristics are strongly associated with water quality, our results identify some less commonly assessed indicators that may influence household water contamination risk. Household animal ownership and observed animal feces – a significant predictor in three of four reduced models – are significantly associated with contamination and could be easily incorporated into household surveys.

This study contributes to a growing body of evidence on the lack of significant associations between payment and water contamination, something inadequately understood.

Limitations

We used an observational, cross-sectional dataset for this analysis. This dataset allowed us to achieve a large sample size across multiple countries, using consistent questions and methods, which are highly costly and pose other feasibility challenges for interventional studies. However, these strengths come with trade-offs and limitations for observational data. The predictive methods used could not generate explanatory models that show causal relationships. Although the household sample size was comparatively large and variance assumptions in regression criterion were met, reducing survey information to meaningful binary or categorical variables reduced the amount of information available for models and thus the power of analyses.

Furthermore, *E. coli* is a measure that can fluctuate across samples at different times of the day or on other days, weeks, months, or seasons of the year; while mitigated by relatively large sample sizes in this study, noise in this measure emphasizes the importance of interpreting these results as point estimates. Cross-sectional sampling limited the ability to make observations across different periods.

Although there are limitations to backward elimination for variable selection, logistic regression modeling for predicting unsafe water contamination is not novel. [24] We narrowed the dataset to a reasonable number of predictor variables using monitored elimination. This was used to

View Article Online
DOI: 10.1039/D3EW00915G

create reduced models with higher context-specific predictive power following the purposes of this research without overly allowing for spurious results from unmonitored modeling of this nature. This approach to creating predictive models reduces many common concerns with elimination methodology, most notably creating "black box" models and conducting spurious variable selection.[27,72,73]

Conclusion

The high prevalence of fecal contamination in household drinking water presents a substantial health risk, and the potential for reducing the water-related disease burden in rural SSA is high. Fecal contamination is prevalent in SSA, especially in less wealthy, less educated, rural households.

Large, observational datasets, such as those we have used in this study, enable the identification of risk factors associated with drinking water contamination across contexts in ways that may be cost- or logistically prohibitive in intervention studies. Our elimination regression analysis showed that season and other well-documented determinants, such as using an unimproved water source, the absence of a water committee, and domestic animal ownership, are significantly associated with household *E. coli* contamination. Some theoretically established determinants – such as household water treatment and storage practices, payment, or sanitation and hygiene – were not found to be significantly related to water contamination.

Findings suggest that policymakers and practitioners should prioritize expanding on-plot water services, support local water source management structures, and incorporate animal husbandry practices into water, sanitation, and hygiene interventions.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

We want to thank the households who participated in this study. We would also like to thank the firms and enumerators who made this study possible and the project sponsor, World Vision, for support, partnerships with in-country experts, and funding of data collection. We want to thank the team at The Water Institute for supporting this work, specifically Amy Guo, Hayley Elling, and Raymond Guo, for their support in reviewing early drafts of this manuscript.

References

1 L. Fewtrell and J. M. Colford, Water, sanitation and hygiene in developing countries: Interventions and diarrhoea-a review, *Water Science and Technology*, 2005, **52**, 133–142.

2 H. Waddington and B. Snilstveit, Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea, *J Dev Effect*, 2009, **1**, 295–335.

3 T. F. Clasen, K. T. Alexander, D. Sinclair, S. Boisson, R. Peletz, H. H. Chang, F. Majorin and S. Cairncross, Interventions to improve water quality for preventing diarrhoea, *Cochrane Database Syst Rev*, 2015, CD004794.

View Article Online
DOI: 10.1039/D3EW00915G

4 R. K. Linsley, Drawers of Water: Domestic Water Use in East Africa, *J Environ Qual*, 1974, **3**, 187–188.

5 Office of the United States Environmental Protection Agency, Drinking Water Contaminants.

6 WHO, *Guidelines for drinking-water quality*, World Health Organization, Geneva, 4th edn., 2011.

7 J. Wolf, R. B. Johnston, A. Ambelu, B. F. Arnold, R. Bain, M. Brauer, J. Brown, B. A. Caruso, T. Clasen, J. M. Colford, J. E. Mills, B. Evans, M. C. Freeman, B. Gordon, G. Kang, C. F. Lanata, K. O. Medlicott, A. Prüss-Ustün, C. Troeger, S. Boisson and O. Cumming, Burden of disease attributable to unsafe drinking water, sanitation, and hygiene in domestic settings: a global analysis for selected adverse health outcomes, *The Lancet*, , DOI:10.1016/S0140-6736(23)00458-0.

8 United Nations, Sustainable Development Goals.

9 WHO/UNICEF, *Progress on household drinking water, sanitation and hygiene 2000-2022: special focus on gender*, United Nations Children's Fund and World Health Organization, New York, 2023.

10 C. A. Osunla and A. I. Okoh, Vibrio pathogens: A public health concern in rural water resources in sub-Saharan Africa, *Int J Environ Res Public Health*, , DOI:10.3390/ijerph14101188.

11 WHO/UNICEF JMP, *Progress on household drinking water, sanitation and hygiene 2000-2017: Special focus on inequalities*, 2019.

12 S. T. Odonkor and T. Mahami, Escherichia coli as a Tool for Disease Risk Assessment of Drinking Water Sources , *Int J Microbiol*, 2020, **2020**, 1–7.

13 R. Bain, R. Cronk, R. Hossain, S. Bonjour, K. Onda, J. Wright, H. Yang, T. Slaymaker, P. Hunter, A. Prüss-Ustün and J. Bartram, Global assessment of exposure to faecal contamination through drinking water based on a systematic review, *Tropical Medicine and International Health*, 2014, **19**, 917–927.

14 A. F. Trevett, R. C. Carter and S. F. Tyrrel, Water quality deterioration: A study of household drinking water quality in rural Honduras, *Int J Environ Health Res*, 2004, **14**, 273–283.

15 K. F. Shields, R. E. S. Bain, R. Cronk, J. A. Wright and J. Bartram, *Environ Health Perspect*, 2015, 123.

16 R. Bain, R. Cronk, J. Wright, H. Yang, T. Slaymaker and J. Bartram, Fecal Contamination of Drinking-Water in Low- and Middle-Income Countries: A Systematic Review and Meta-Analysis, *PLoS Med*, , DOI:10.1371/journal.pmed.1001644.

17 K. F. Shields, R. E. S. Bain, R. Cronk, J. A. Wright and J. Bartram, Association of Supply Type with Fecal Contamination of Source Water and Household Stored Drinking Water in Developing Countries: A Bivariate Meta-analysis, *Environ Health Perspect*, 2015, **123**, 1222–1231.

18 A. Ercumen, A. J. Pickering, L. H. Kwong, B. F. Arnold, S. M. Parvez, M. Alam, D. Sen, S. Islam, C. Kullmann, C. Chase, R. Ahmed, L. Unicomb, S. P. Luby and J. M. Colford, Animal Feces Contribute to Domestic Fecal Contamination: Evidence from E. coli Measured in Water, Hands, Food, Flies, and Soil in Bangladesh, *Environ Sci Technol*, , DOI:10.1021/acs.est.7b01710.

19 R. Bain, R. Johnston, S. Khan, A. Hancioglu and T. Slaymaker, Monitoring drinking water quality in nationally representative household surveys in low-and middle-income countries: Cross-sectional analysis of 27 multiple indicator cluster surveys 2014-2020, *Environ Health Perspect*, , DOI:10.1289/EHP8459.

20 M. A. Kirby, C. L. Nagel, G. Rosa, L. Iyakaremye, L. D. Zambrano and T. F. Clasen, Faecal contamination of household drinking water in Rwanda: A national cross-sectional study, *Science of the Total Environment*, 2016, **571**, 426–434.

21 The Water Institute at UNC and World Vision Inc., *Household Deidentified Survey Data*, UNC Dataverse, V1, 2020.

22 The Water Institute at UNC, *The World Vision 14-Country Evaluation Final Report.*, 2019.

23 C. Stauber, C. Miller, B. Cantrell and K. Kroell, Evaluation of the compartment bag test for the detection of Escherichia coli in water, *J Microbiol Methods*, 2014, **99**, 66–70.

View Article Online
DOI: 10.1039/D3EW00915G

24 S. T. McGarvey, J. Buszin, H. Reed, D. C. Smith, Z. Rahman, C. Andrzejewski, K. Awusabo-Asare and M. J. White, Community and household determinants of water quality in coastal Ghana, *J Water Health*, , DOI:10.2166/wh.2008.057.

25 WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation (JMP), JMP 2018 Core Questions for Household Surveys.

26 K. Dietz, M. Gail, K. Krickeberg, J. Samet, A. Tsiatis, T. M. Therneau and P. M. Grambsch, *Statistics for Biology and Health*, 2000.

27 E. Suárez, C. M. Pérez, R. Rivera and M. N. Martínez, *Applications of Regression Models in Epidemiology*, 2017.

28 A. Cassivi, S. Guilherme, R. Bain, E. Tilley, E. O. D. Waygood and C. Dorea, Drinking water accessibility and quantity in low and middle-income countries: A systematic review, *Int J Hyg Environ Health*, 2019, **222**, 1011–1020.

29 E. Kumpel, R. Peletz, M. Bonham and R. Khush, Assessing Drinking Water Quality and Water Safety Management in Sub-Saharan Africa Using Regulated Monitoring Data, *Environ Sci Technol*, 2016, **50**, 10869–10876.

30 K. Onda, J. LoBuglio and J. Bartram, Global Access to Safe Water: Accounting for Water Quality and the Resulting Impact on MDG Progress, *Int J Environ Res Public Health*, 2012, **9**, 880–894.

31 S. Godfrey, P. Labhasetwar, S. Wate and S. Pimpalkar, How safe are the global water coverage figures? Case study from Madhya Pradesh, India, *Environ Monit Assess*, 2011, **176**, 561–574.

32 R. Bain, S. W. Gundry, J. A. Wright, H. Yang, S. Pedley and J. K. Bartram, Accounting for water quality in monitoring the Millennium Development Goal on access to safe drinking-water : lessons from five countries, *Bull World Health Organ*, , DOI:10.2471/BLT.11.094284.

33 M. A. Usman, N. Gerber and E. H. Pangaribowo, Drivers of microbiological quality of household drinking water – a case study in rural Ethiopia, *J Water Health*, , DOI:10.2166/wh.2017.069.

34 E. Kelly, K. F. Shields, R. Cronk, K. Lee, N. Behnke, T. Klug and J. Bartram, Seasonality, water use and community management of water systems in rural settings: Qualitative evidence from Ghana, Kenya, and Zambia, *Science of the Total Environment*, 2018, **628–629**, 715–721.

35 P. Jagals, Does improved access to water supply by rural households enhance the concept of safe water at the point of use? A case study from deep rural South Africa, *Water Science and Technology*, 2006, **54**, 9–16.

36 K. Sakisaka, E. A. Chadeka, S. Nagi, D. S. Mwandembo and M. Jimba, Introduction of a community water supply in rural western Kenya: Impact on community wellbeing and child health, *Int Health*, 2015, **7**, 204–211.

37 T. Klug, K. F. Shields, R. Cronk, E. Kelly, N. Behnke, K. Lee and J. Bartram, Water system hardware and management rehabilitation: Qualitative evidence from Ghana, Kenya, and Zambia, *Int J Hyg Environ Health*, 2017, **220**, 531–538.

38 N. L. Behnke, T. Klug, R. Cronk, K. F. Shields, K. Lee, E. R. Kelly, G. Allgood and J. Bartram, Resource mobilization for community-managed rural water systems: Evidence from Ghana, Kenya, and Zambia, *J Clean Prod*, , DOI:10.1016/j.jclepro.2017.04.016.

39 L. Whaley and F. Cleaver, Can 'functionality' save the community management model of rural water supply?, *Water Resources and Rural Development*, 2017, **9**, 56–66.

40 P. Hutchings, M. Y. Chan, L. Cuadrado, F. Ezbakhe, B. Mesa, C. Tamekawa and R. Franceys, A systematic review of success factors in the community management of rural water supplies over the past 30 years, *Water Policy*, 2015, **17**, 963–983.

41 E. Kelly, K. Lee, K. F. Shields, R. Cronk, N. Behnke, T. Klug and J. Bartram, The role of social capital and sense of ownership in rural community-managed water systems: Qualitative evidence from Ghana, Kenya, and Zambia, *J Rural Stud*, 2017, **56**, 156–166.

View Article Online
DOI: 10.1039/D3EW00915G

42 M. B. Fisher, K. F. Shields, T. U. Chan, E. Christenson, R. D. Cronk, H. Leker, D. Samani, P. Apoya, A. Lutz and J. Bartram, *Water Resour Res*, 2015.

43 M. N. B. Momba and P. Kaleni, Regrowth and survival of indicator microorganisms on the surfaces of household containers used for the storage of drinking water in rural communities of South Africa, *Water Res*, 2002, **36**, 3023–3028.

44 P. Jagals, C. Jagals and T. C. Bokako, The effect of container-biofilm on the microbiological quality of water used from plastic household containers, *J Water Health*, 2003, **1**, 101–108.

45 E. J. Lee and K. J. Schwab, Deficiencies in drinking water distribution systems in developing countries, *J Water Health*, 2005, **3**, 109–127.

46 G. Rosa and T. Clasen, Estimating the scope of household water treatment in low- and medium-income countries, *American Journal of Tropical Medicine and Hygiene*, 2010, **82**, 289–300.

47 A. F. Trevett, R. C. Carter and S. F. Tyrrel, The importance of domestic water quality management in the context of faecal-oral disease transmission, *J Water Health*, 2005, **3**, 259–270.

48 T. F. Clasen and A. Bastable, Faecal contamination of drinking water during collection and household storage: The need to extend protection to the point of use, *J Water Health*, , DOI:10.2166/wh.2003.0013.

49 D. M. Anderson, M. B. Fisher, O. Kwena, H. Kambou, R. Broseus, A. R. Williams, K. Liang, R. Ramaswamy and J. Bartram, Adapting a safe water storage container to improve household stored water quality in rural Burkina Faso: a cluster randomized trial, *Journal of Water, Sanitation and Hygiene for Development*, 2021, **11**, 719–731.

50 M. B. Fisher, L. Danquah, Z. Seidu, A. N. Fechter, B. Saga, J. K. Bartram, K. M. Liang and R. Ramaswamy, WaSH CQI: Applying continuous quality improvement methods to water service delivery in four districts of rural northern Ghana, *PLoS One*, 2020, **15**, e0233679.

51 J. Wright, S. Gundry and R. Conroy, Household drinking water in developing countries: A systematic review of microbiological contamination between source and point-of-use, *Tropical Medicine and International Health*, 2004, **9**, 106–117.

52 J. Brown and M. Sobsey, Evaluating household water treatment options: Heath-based targets and microbiological performance specifications, *World Health Organization (WHO) Publications*, 2011, 68.

53 K. Levy, Does Poor Water Quality Cause Diarrheal Disease?, *Am J Trop Med Hyg*, 2015, **93**, 899.

54 J. Tucker, A. MacDonald, L. Coulter and R. C. Calow, Household water use, poverty and seasonality: Wealth effects, labour constraints, and minimal consumption in Ethiopia, *Water Resources and Rural Development*, 2014, **3**, 27–47.

55 O. M. Morakinyo, S. A. Adebowale and E. O. Oloruntoba, Wealth status and sex differential of household head: Implication for source of drinking water in Nigeria, *Archives of Public Health*, 2015, **73**, 1–9.

56 R. Hope and P. Ballon, Global water policy and local payment choices in rural Africa, , DOI:10.1038/s41545-019-0045-y.

57 N. A. Wardrop, A. G. Hill, M. Dzodzomenyo, G. Aryeetey and J. A. Wright, Livestock ownership and microbial contamination of drinking-water: Evidence from nationally representative household surveys in Ghana, Nepal and Bangladesh, *Int J Hyg Environ Health*, 2018, **221**, 33–40.

58 A. N. Barnes, J. D. Anderson, J. Mumma, Z. H. Mahmud and O. Cumming, The association between domestic animal presence and ownership and household drinking water contamination among peri-urban communities of Kisumu, Kenya, *PLoS One*, 2018, **13**, e0197587.

59 K. L. Kotloff, J. P. Nataro, W. C. Blackwelder, D. Nasrin, T. H. Farag, S. Panchalingam, Y. Wu, S. O. Sow, D. Sur, R. F. Breiman, A. S. G. Faruque, A. K. M. Zaidi, D. Saha, P. L. Alonso, B. Tamboura, D. Sanogo, U. Onwuchekwa, B. Manna, T. Ramamurthy, S. Kanungo, J. B. Ochieng, R. Omore, J. O. Oundo, A. Hossain, S. K. Das, S. Ahmed, S. Qureshi, F. Quadri, R. A. Adegbola, M. Antonio, M. J.

View Article Online
DOI: 10.1039/D3EW00915G

Hossain, A. Akinsola, I. Mandomando, T. Nhampossa, S. Acácio, K. Biswas, C. E. O'Reilly, E. D. Mintz, L. Y. Berkeley, K. Muhsen, H. Sommerfelt, R. M. Robins-Browne and M. M. Levine, Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): A prospective, case-control study, *The Lancet*, 2013, **382**, 209–222.

60 C. Kostyla, R. Bain, R. Cronk and J. Bartram, *Science of the Total Environment*, 2015, 514, 333–343.

61 G. Howard, S. Pedley, M. Barrett, M. Nalubega and K. Johal, Risk factors contributing to microbiological contamination of shallow groundwater in Kampala, Uganda, *Water Res*, 2003, **37**, 3421–3429.

62 A. A. Cronin, N. Breslin, J. Gibson and S. Pedley, Monitoring source and domestic water quality in parallel with sanitary risk identification in Northern Mozambique to prioritise protection interventions, *J Water Health*, 2006, **4**, 333–345.

63 J. N. Edokpayi, E. T. Rogawski, D. M. Kahler, C. L. Hill, C. Reynolds, E. Nyathi, J. A. Smith, J. O. Odiyo, A. Samie, P. Bessong and R. Dillingham, Challenges to sustainable safe drinking water: A case study ofwater quality and use across seasons in rural communities in Limpopo Province, South Africa, *Water (Switzerland)*, , DOI:10.3390/w10020159.

64 United Nations Children's Fund and World Health Organization, *Integrating Water Quality Testing into Household Surveys: Thematic report on drinking water*, UNICEF and WHO, New York, 2020.

65 J. Brown, V. T. Hien, L. McMahan, M. W. Jenkins, L. Thie, K. Liang, E. Printy and M. D. Sobsey, Relative benefits of on-plot water supply over other 'improved' sources in rural Vietnam, *Tropical Medicine & International Health*, 2013, **18**, 65–74.

66 G. Howard and J. Bartram, *Sanitation and Health Programme*, .

67 C. L. Moe and R. D. Rheingans, Global challenges in water, sanitation and health, *J Water Health*, , DOI:10.2166/wh.2006.0043.

68 G. Hutton and L. Haller, *Evaluation of the Costs and Benefits of Water and Sanitation Improvements at the Global Level*, 1994.

69 S. Budge, A. Ambelu, J. Bartram, J. Brown and P. Hutchings, Environmental sanitation and the evolution of water, sanitation and hygiene, *Bull World Health Organ*, 2022, **100**, 286–288.

70 G. Penakalapati, J. Swarthout, M. J. Delahoy, L. McAliley, B. Wodnik, K. Levy and M. C. Freeman, Exposure to Animal Feces and Human Health: A Systematic Review and Proposed Research Priorities, *Environ Sci Technol*, 2017, **51**, 11537–11552.

71 WHO, One Health.

72 G. Shmueli, To explain or to predict?, *Statistical Science*, 2010, **25**, 289–310.

73 P. C. Austin and J. V. Tu, Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality, *J Clin Epidemiol*, 2004, **57**, 1138–1146.