

## ARTICLE OPEN



# Sanitary inspection, microbial water quality analysis, and water safety in handpumps in rural sub-Saharan Africa

Emma Kelly<sup>1</sup>✉, Ryan Cronk<sup>1,2</sup>, Michael Fisher<sup>1</sup> and Jamie Bartram<sup>1,3</sup>

In sub-Saharan Africa, over half of the population is exposed to contaminated drinking water. The WHO recommends both sanitary inspection and water quality analysis to assess the risk of water source contamination, but the relationship between these tools is poorly understood. We explore the relationship between sanitary inspection and water quality analysis using data from 1028 boreholes with handpumps in 12 countries in sub-Saharan Africa. Sanitary inspection scores and *E. coli* occurrence were compared using the models described in published literature, and an alternative model that better reflects causal pathways of contamination. In the alternative model, sanitary risk factors were categorized as contamination sources, carriers, or barrier breakdowns, and the relationships between risk factor combinations and *E. coli* occurrence were assessed. We found no associations between sanitary risk score and *E. coli* occurrence using either the established or alternative model. These results confirm that sanitary inspections and microbial analyses convey distinct information, and perfect correlation is neither expected nor desired. The alternative model demonstrated a slightly better model fit than most established models, and the model fit further improved when the occurrence of rainfall in the past two days was added as a carrier. We recommend that: implementers train water system operators to conduct sanitary inspection; and researchers work to improve our understanding of the effect of individual sanitary risk factors, as well as incorporate contextual data into their assessments of sanitary inspection and water quality.

npj Clean Water (2021)4:3; <https://doi.org/10.1038/s41545-020-00093-z>

## INTRODUCTION

In sub-Saharan Africa (SSA) 42% of people lack access to a basic water supply, defined as an improved water source accessible within a 30 min fetching time<sup>1</sup>. However, even people using a basic water supply may still be drinking contaminated water, as water quality is not considered directly. While improved water sources generally tend to produce water with better microbial quality than unimproved supplies, a systematic review shows that many improved sources are contaminated<sup>2</sup>. One 2012 study estimated that 52% of the SSA population was exposed to fecally contaminated drinking water<sup>2</sup>, and another reported that 31% of boreholes/tubewells in SSA exceeded the World Health Organization (WHO) guideline of no detectable fecal indicator bacteria per 100-mL sample<sup>3</sup>. Ensuring water safety is especially difficult in rural SSA, where many people rely on small, community-managed systems for drinking water. These rural systems tend to have even poorer water quality than urban water systems<sup>4,5</sup>.

Risk-based preventative management is recommended to protect and improve water safety in all settings. As the WHO suggests, “the most effective means of consistently ensuring the safety of a drinking-water supply is through the use of a comprehensive risk assessment and risk management approach”<sup>6</sup>. Sanitary inspection is a visual survey of risk factors that may contribute to the likelihood of fecal contamination in water systems, and is considered an effective and low-cost tool for risk assessment<sup>7</sup>. In small community-managed water systems, sanitary inspections are typically conducted using forms developed by the WHO, or adaptations of these<sup>8</sup>. The forms are technology-type specific, and each comprises 9–12 yes/no questions representing the presence/absence of sanitary risk factors. The total number of sanitary risk factors is summed to an

overall sanitary risk score, which is often used to compare the level of risk between systems.

Several studies have explored the relationship between sanitary inspection and water quality. Although sanitary inspection and water quality are conceptually linked<sup>9</sup>, some studies found no significant association between the two<sup>10–13</sup>. Water supply professionals have questioned the effectiveness of sanitary inspection because of this apparent contradiction<sup>13</sup>. However, previous assessments of the relationship have common weaknesses<sup>14</sup>. Literature comparing sanitary inspection and water quality analysis has depended on the untested assumption that sanitary risk score and *E. coli* concentration of a water source are positively and linearly associated in all cases. This has been assessed by applying well-known statistical tests to compare *E. coli* concentration (as a continuous, ordinal, or presence/absence measure) to sanitary risk scores. Sanitary risk score, however, is not a comprehensive representation of system risk at a given point in time but is a simplified output of a tool designed to identify observable risk factors and guide corrective action<sup>14</sup>.

This study explores the nature of the relationship between sanitary inspection and microbial water quality analysis from a new perspective, in order to improve the general understanding of each tool and identify opportunities to enhance the effective application of each. We address the following questions:

1. What is the nature of the relationship between sanitary risk score and *E. coli* occurrence in water supplies?
2. How best can this relationship be quantified and modeled?
3. Can delivery and monitoring of drinking-water service be improved based on these findings?

We used data collected from 1028 boreholes with handpumps in rural SSA to explore the relationship between sanitary

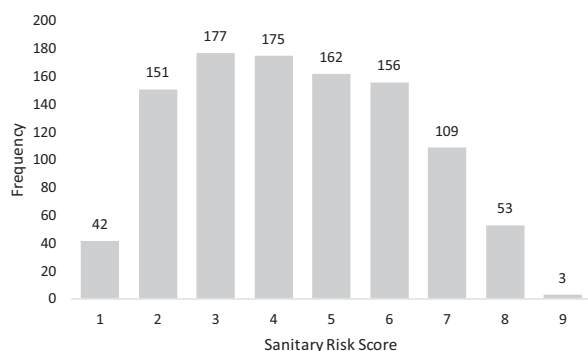
<sup>1</sup>The Water Institute, University of North Carolina, Chapel Hill, NC, USA. <sup>2</sup>ICF, 2635 Meridian Pkwy Suite 200, Durham, NC 27713, USA. <sup>3</sup>School of Civil Engineering, University of Leeds, Leeds, UK. ✉email: [erkelly@alumni.unc.edu](mailto:erkelly@alumni.unc.edu)

inspection and microbial water quality. The data were analyzed using the statistical models applied in previous studies, and using an alternative statistical model that, we suggest, more accurately represents the causal pathways of water system contamination. We compared these models to evaluate model fit and considered implications with respect to our research questions.

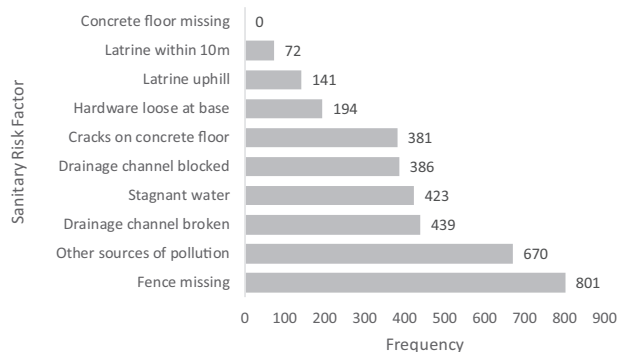
## RESULTS

### Descriptive statistics

Of 1028 handpumps included in our sample, 805 (78%) had no detectable *E. coli* in a 100-mL sample. Sanitary risk scores ranged from zero (low risk) to nine (high risk), with an average sanitary risk score of 3.4 (Fig. 1). The sanitary risk scores were normally distributed.



**Fig. 1** Frequency of sanitary risk scores at 1028 handpumps with boreholes in sub-Saharan Africa. The frequency of sanitary risks scores shows that 2-7 sanitary risk factors were identified at most handpumps and zero received a risk score of 10, which indicates the highest levels of sanitary risk.



**Fig. 2** Frequency of individual sanitary risk factors at 1028 handpumps in sub-Saharan Africa. The number of handpumps at which each risk included on the WHO borehole with handpump sanitary inspection form was present.

The number of handpumps where each sanitary risk factor was present is shown in Fig. 2. A missing fence was the most commonly observed risk factor (78% of handpumps) and the presence of a latrine within 10 m was the least common risk factor (1% of handpumps) (Fig. 2).

### Established models

Sanitary inspection results and *E. coli* occurrence from the 1028 handpumps were compared using previously published statistical models (Table 1). No model showed a significant association between *E. coli* occurrence and either sanitary risk score or sanitary risk score class at a 95% confidence level. Of the models using a sanitary risk score or sanitary risk class as the dependent variable, logistic regression with a binary *E. coli* presence/absence variable demonstrated the best model fit (AIC = 1079).

Previously published models also used multivariable regressions to assess the relationship between individual sanitary risk factors and *E. coli* occurrence. Logistic regression comparing a binary presence/absence *E. coli* dependent variable and individual sanitary risk factors had the best model fit (AIC = 1074) (results in Table 2). Logistic regression revealed a significant association between *E. coli* presence/absence and several individual sanitary risk factors. The presence of a latrine on higher ground (OR = 0.60,  $p = 0.046$ ), a broken drainage channel (OR = 1.8,  $p = 0.005$ ), and cracks in the apron (OR = 0.56,  $p = 0.002$ ) were significantly associated with *E. coli* presence at the 95% confidence level (Table 2).

### Alternative model

Table 3 shows the distribution of the source, carrier, and barrier breakdown risk scenarios. The largest proportion of handpumps (41%,  $n = 421$ ) had a contamination source, carrier, and barrier breakdown present at the time of data collection. Only 2% ( $n = 24$ ) handpumps had no sanitary risk factors.

The results of logistic regression comparing the *E. coli* presence/absence variable and each of the sanitary risk scenarios are shown in Table 3. One scenario was significantly associated with contamination at the 95% confidence level—where only the breakdown is present and the source and carrier are absent. Two more scenarios were significantly associated with *E. coli* occurrence at a 90% confidence level, including the expected scenario in which a source, carrier, and barrier are present. The alternative model had an AIC value of 1079 and  $R^2 = 0.0086$ .

The alternative model initially had one contamination carrier (stagnant water within 2 m). When rainfall within the past 2 days was added as another potential carrier, the AIC decreased to 1070 and the  $R^2$  increased to 0.019. The results of the model in which rainfall is included are shown in Table 3. When rainfall is included, none of the risk scenarios are significantly associated with *E. coli* occurrence at the 95% confidence level.

**Table 1.** Results of previous models showing no significant association between sanitary risk score and water quality analysis—AIC values show that logistic regression is the model with the best fit.

Model	Independent variable	Dependent variable	$p > t$	$R^2$	AIC
Logistic regression	Present/absent	Sanitary risk score	0.62	0.00020	1079
Ordinal log regression	WHO risk category	Sanitary risk score	0.45	0.00040	1557
Pearson's chi-square	WHO risk category	Sanitary risk class	0.45		1557
Logistic regression	Present/absent	Sanitary risk factors		0.021	1074
Ordinal log regression	WHO risk category	Sanitary risk factors		0.014	1553

## DISCUSSION

This study explored the relationship between sanitary inspection and microbial water quality using data from more than 1000 handpump across 12 countries in SSA. The results validate the use of an alternative model, which reflects the causal pathway of water system contamination, and shows that the relationship between sanitary inspection and microbial contamination can be modeled more thoughtfully. When the previous statistical models were applied, they revealed no significant correlation between sanitary risk score and *E. coli* occurrence. When our alternative model—which accounts for relationships between contamination sources, carriers, and barrier breakdowns—was applied, it demonstrated a weak but significant association between handpump breakdowns and *E. coli* occurrence. Furthermore, it demonstrates that some sanitary risk factors, especially barrier breakdowns, are more strongly associated with water quality than others.

The alternative model revealed that the risk scenario in which a barrier breakdown is present, but the contamination source and carrier absent was the only scenario significantly associated with *E. coli* occurrence (OR = 3.6,  $p = 0.019$ ). This may suggest that barrier breakdown is the best predictor of *E. coli* occurrence, that important contamination sources and carriers are missing from the current sanitary inspection forms and models, and/or that sources and carriers are more difficult to reliably capture using sanitary inspection tools than barrier breakdowns. The limited role of carrier risk factors in the model is perhaps unsurprising since only one sanitary risk factor (stagnant water within two meters) falls into the carrier category. In order to address this, recent rainfall—previously reported to be an important predictor and carrier of microbial contamination<sup>10,13,15</sup>—was added to the

model. When rainfall was incorporated as a carrier of contamination, the fit of the alternative model improved slightly (AIC: 1079 to 1070,  $R^2$ : 0.0086 to 0.019), but none of the risk scenarios were significantly associated with *E. coli* occurrence.

It is not surprising that statistical models comparing sanitary risk score and *E. coli* occurrence produce no significant association, in line with the results reported by the studies that used the established models<sup>12,13,16–18</sup>. Sanitary inspection was designed as a risk assessment and decision-making tool, and is expected to be only loosely correlated with *E. coli* occurrence in a given water sample at a given point in time<sup>14</sup>. This loose correlation is demonstrated by the  $R^2$  values, which did not exceed 3% in any of the models in this study. We suggest three explanations for the small  $R^2$  values: (1) a single *E. coli* measurement does not accurately represent the microbial water quality of a given water system over time; (2) current WHO sanitary inspection forms do not adequately identify all relevant sources, carriers and barrier breakdowns affecting water system contamination; (3) the use of a model that does not weight risk factors distorts associations in which a small number of important factors dominate microbial risk. It is possible that one, two, or all three of these explanations is applicable in some settings.

A single microbial water sample is unlikely to accurately represent the central tendency of water quality in a water system. *E. coli* occurrence and concentration are highly variable over space and time. Previous studies have shown that microbes are not homogeneously distributed in water supplies, meaning that two water quality samples could contain very different *E. coli* concentrations even if collected from the same source at the same time<sup>19</sup>. Microbial contamination also varies dramatically over short time periods, and is influenced by seasonal weather patterns and recent rainfall<sup>20</sup>.

The alternative model was designed based on the assumption that a contamination source, carrier, and barrier breakdown would all be necessary for contamination to occur at the water system. However, the results show that only the scenario significantly associated with *E. coli* occurrence is the one in which a barrier breakdown alone is present. This suggests that the sanitary risk factors in current forms may not adequately capture all relevant contamination sources, carriers, and barrier breakdowns for water systems in the study setting. Environmental, technical, and social factors not included in current sanitary inspection forms may also be important in explaining contamination. Topography<sup>15</sup>, water system age<sup>21</sup>, and population density<sup>22</sup>, for example, have all been identified as determinants of water system contamination, and are not incorporated in the WHO sanitary inspection forms.

Finally, the WHO sanitary inspection forms do not weight risk factors, but some factors may be more influential than others. The presence of cracks in a slab, for example, maybe more or less strongly associated with microbial contamination than the

**Table 2.** Results of logistic regression assessment of individual sanitary risk factors—sanitary risk factors with \* are significantly associated with *E. coli* presence/absence at the 95% confidence level.

Variables	Odds ratio (OR)	$p > t$
Latrine within 10 m	1.0	0.92
Latrine on higher ground	0.60	0.046*
Other pollution source within 10 m	1.0	0.81
Drainage channel blocked	1.1	0.63
Drainage channel broken	1.8	0.005*
Fence missing	1.2	0.28
Incomplete/missing concrete apron	0.87	0.47
Stagnant water	0.80	0.19
Cracks in apron	0.56	0.002*
Hardware loose at base	1.0	0.82

**Table 3.** Summary statistics describing possible risk scenarios, with logistic regression results for the alternative model both with and without rainfall, incorporated—risk scenarios with \* are significantly associated with *E. coli* presence/absence at the 95% confidence level.

Risk scenario						Without rainfall		With rainfall	
Source	Carrier	Breakdown	$n$	% of Total	% Low risk	OR	$p > t$	OR	$p > t$
X	X	X	421	41	75	2.6	0.076	3.6	0.086
X		X	225	21	84	2.7	0.069	2.2	0.31
X			42	4	90	2.2	0.22	1.2	0.87
X	X		29	2	69	2	0.39	5	0.057
		X	131	13	80	3.6	0.019*	2.7	0.19
	X	X	136	13	71	2.1	0.23	4.4	0.051
	X		20	1	90	1	Omitted	1.2	0.85
			24	2	92				

absence of a fence, and the relative importance of different factors may even vary by setting, season, and/or construction quality. In many cases, one serious barrier breakdown may dominate all other risk factors in that category. Such limitations might reduce the apparent association between sanitary inspection and microbial water quality. However, all sanitary risk factors would presumably be worth noting and addressing, and thus the current sanitary inspection approach may be considered more suitable as a risk assessment tool than as a proxy for microbial data.

The results of this study have implications for rural water service management, policy, and further research. With regard to management, sanitary inspection should be prioritized over water quality analysis as a means of identifying and prioritizing repairs to water systems. It is especially valuable in settings where water quality testing is infrequent or infeasible. Although sanitary inspection and water quality analysis are recommended as complementary tools, many small rural water systems in SSA do not conduct water quality monitoring with the frequency recommended by WHO<sup>23</sup> or others<sup>24</sup>. Therefore, a single test that fails to detect *E. coli* may give managers a false sense of security if not accompanied by sanitary inspection data. Sanitary inspection results vary less over time than microbial water quality, require less technical training to perform, and provide actionable information even when carried out infrequently. The operator of even a small water system should be able to conduct a sanitary inspection independently one or more times per year; operators should be trained, or at least invited in a routine inspection carried out by governments and other implementers in order to learn. When a sanitary risk factor is identified, action should be taken to repair the system or otherwise mitigate the risk. Corrective actions might include sealing cracks and leaks, repairing or replacing damaged or absent components, or, in extreme cases, rehabilitating damaged systems. Operators should be trained to carry out basic remedial actions, or access external support for more involved remediation.

Implications for policy include potential modifications to current sanitary inspection forms to enhance their utility for assessing water system vulnerabilities. Additional evidence-based risk carriers and sources may be included in sanitary inspections, or open access data can be downloaded and incorporated after sanitary inspection data collection, as appropriate.

Implications for research include the need to develop larger and richer datasets linking sanitary inspection data, longitudinal meteorological, water quality, and/or health outcome data, and to use these datasets to further validate and improve current sanitary inspection tools and forms. Such activities may include testing the integration of season as a variable or covariate in sanitary inspections, investigating the relevance of additional risk factors (such as well depth, intermittent versus continuous water supply, soil characteristics, community factors, etc.) in sanitary inspections, and/or further exploring the importance of interactions among sanitary risk factors. Exploration of the extent to which the importance of different sanitary risk factors varies across different settings could also inform further refinements.

There are several limitations to this work. Although a sample size of 1028 was sufficient for regression analysis, the dataset aggregates data from 12 geographically diverse countries in SSA. A larger dataset would have enabled disaggregation at the country level, which might have yielded country- and region-specific findings. Models used in this study were relatively simple and did not include water system variables such as well depth, soil characteristics, continuous versus intermittent pressure of piped systems, etc., or control for covariates such as community size, degree of rurality, population density, subsurface geology, the intensity of animal agriculture activities, land use and sanitation variables, and other factors capable of contributing to fecal contamination of water systems; including such variables in future

analyses may improve associations between sanitary inspection scores and microbial water quality data.

With respect to the alternative model used in this study, some risk factors were difficult to classify. While the presence of damaged or cracked drainage channels was classified as a barrier factor, this could have been classified as a carrier factor, since it relates to the release of pumped water which can infiltrate the borehole and transport nearby contamination into the water supply.

Our updated alternative model provides a simple conceptual framework for representing and evaluating sanitary risk factors with respect to their likely function in microbial transmission pathways. In the current work, each risk category (contamination source, carrier, barrier breakdown) is classified as either present or absent. A more detailed method that both incorporate these groups of factors and the individual risk factors within them, and/or accounts for the cumulative impact of multiple risk factors in a given category, may enhance the value of these sanitary inspection tools. As noted above, the benefits of any such improvements with respect to diagnostic and predictive power would have to be weighed against any associated loss of simplicity before determining whether their inclusion is appropriate for routine sanitary inspections.

Sanitary inspection and water quality analysis are complementary, valuable tools for characterizing and protecting water safety. Sanitary inspection is especially useful in settings where water quality analysis is not carried out frequently, as it enables water safety risks to be identified and addressed. As service levels in rural SSA increase, we must consider the most effective ways to protect water safety in rural water systems, and the relevance of sanitary inspection as a risk assessment tool is likely to be enhanced. We recommend that water system operators be trained in sanitary inspection and remedial action, and conduct inspections regularly. Furthermore, sanitary inspections may more effectively lead to timely and appropriate action if they are combined with mobile survey tools and/or semi-automated analysis tools to aid rapid response and effective decision-making. As these monitoring and implementation efforts are undertaken, researchers should also seek to continue to collect and compile data on microbial water quality, sanitary risk, and relevant outcomes and to continue to explore the relationship between water system vulnerability and contamination. Such exploration will benefit from using a framework that incorporates evidence on the causal pathways of contamination, as well as social and environmental determinants of contamination, and that facilitates timely and effective response to identified risks to water safety.

## METHODS

### Study population

Data were collected as part of an evaluation of water, sanitation, and hygiene (WaSH) programs implemented by the international non-governmental organization (NGO) World Vision (WV). The data were collected from rural communities in 14 countries—Ethiopia, Ghana, Honduras, India, Kenya, Malawi, Mali, Mozambique, Niger, Rwanda, Tanzania, Uganda, Zambia, and Zimbabwe. For this analysis, Honduras and India were excluded because they are the only countries that do not fall in the sub-Saharan Africa SDG region or the lowest access WaSH country cluster<sup>25</sup>.

### Study design

In-country enumerators and research supervisors were hired in each country to carry out data collection. Research supervisors were trained by staff from The Water Institute at UNC; enumerators were trained by research supervisors.

In each country, a list of rural subnational administrative sampling units (e.g., blocks, districts) was obtained from either the national census bureau or national statistics office. Subnational administrative sampling units were stratified into World Vision and comparison areas, based on whether World Vision worked in the district. For each country, 56 WV and 56 comparison

primary sampling units (PSUs) were then randomly selected from the list of subnational administrative sampling units using a probability proportional to size method. If the PSU was estimated to include more than 200 households, it was divided into secondary sampling (SSU) with 200 households or less, and one SSU was randomly selected to represent that PSU.

Once the 112 sampling units were selected, trained enumerators mapped all households and water sources within the community. In each sampling unit, 25 households were randomly selected for a household survey. All unimproved and improved water sources were surveyed. All water source surveys included a sanitary inspection.

Up to five functional, non-surface water sources were randomly selected for water quality analysis in each sampling unit. 'Functional' systems were defined as those producing water at the time of data collection. If there were fewer than five water sources meeting these criteria in a given sampling unit, all eligible water sources in the sampling unit underwent water quality analysis.

Only boreholes with handpumps (referred to as "handpumps" from here on) were included in this analysis in order to ensure comparability between water sources.

### Sanitary inspection

Sanitary inspections were conducted using an adapted version of the WHO sanitary inspection form for tubewells (boreholes) with handpumps<sup>8</sup>. The form comprises ten yes/no questions listed in Supplementary Table 1. Each water source is evaluated using the form and receives a sanitary risk score ranging from zero to ten, where zero indicates that none of the evaluated sanitary risk factors are present at the handpump and a ten indicates that all are present. Handpumps are categorized into sanitary risk classes similar to those used in other studies<sup>9,26,27</sup>: low risk (0–2), intermediate-risk (3–5), high risk (6–8), and very high risk (9–10).

### Water quality analysis

Samples from the selected water sources were tested to determine *E. coli* concentration (most probable number [MPN] in a 100-mL sample) using the Compartment Bag Test (CBT) field test (Aquagenx LLC, Chapel Hill, NC, USA)<sup>28</sup>. Briefly: samples were collected from the selected handpumps using 100-mL WhirlPak bags and processed immediately according to the manufacturer's instructions. Processed CBTs were incubated at ambient temperature for 48 h where temperatures were between 25 and 30 °C, or for 24 h where temperatures were above 30 °C. Where ambient temperatures were below 25 °C, thermal incubation at 35–37 °C for 24 h was used (field incubator). Results were recorded as *E. coli* MPN/100 mL and categorized into an ordinal variable based on WHO health risk categories and a binary "presence" ( $\geq 1$  *E. coli* MPN/100 mL) or "absence" ( $< 1$  *E. coli* MPN/100 mL) variable<sup>29</sup>.

Field blanks and duplicate field samples were collected from a random 10% of water sources and analyzed similarly. For field blanks, an unopened bottle from a reputable local brand of bottled water was opened at the water source and 100 mL of the bottled water was processed. Duplicate water samples were collected immediately after the collection of the experimental sample. All field blanks and duplicates were processed at the same time and in the same location as the corresponding experimental sample, using identical procedures. All samples (experimental, duplicate, and blank) were labeled with barcodes, reducing the likelihood that enumerators might introduce bias due to knowledge of the sample source when reading CBT results.

### Established models for statistical analysis

Statistical analyses were carried out in Stata/SE 13.1 (StataCorp). Summary statistics were calculated to describe the distribution of *E. coli* MPN/100 mL and sanitary inspection score results.

Statistical methods described in previous literature<sup>14</sup> were used to assess the relationship between sanitary inspection and *E. coli* occurrence:

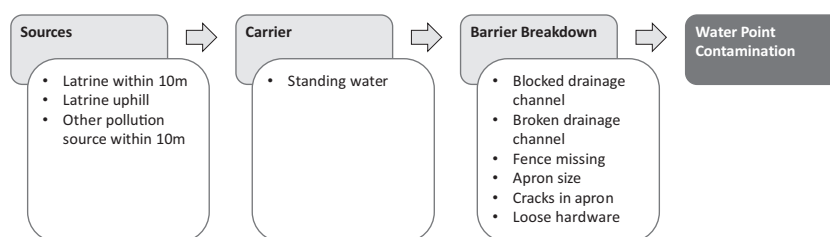
- Logistic regression: used to analyze the relationship between the presence/absence of detectable *E. coli* in 100-mL sample (binary) and sanitary risk score (ordinal), and to compare presence/absence of detectable *E. coli* (binary) with the presence/absence of individual sanitary risk factors (binary)<sup>18,27,30,31</sup>.
- Ordinal logistic regression: used to compare the WHO water quality risk category (ordinal) with both sanitary risk score (ordinal) and the presence/absence of individual sanitary risk factors (binary).
- Non-parametric tests: Pearson's chi-square test was used to determine whether *E. coli* presence/absence (binary) was independent of sanitary risk score (ordinal)<sup>15,32</sup>.

### Alternative model for statistical analysis

The sanitary risk factors included on the WHO borehole with handpump form represent contamination sources, carriers, and barrier breakdowns<sup>14</sup>, defined as reservoirs of pathogens such as nearby latrines or fertilized fields; vehicles that transport pathogens from sources into the system such as surface runoff; and weaknesses in system infrastructure, such cracks in the concrete floor or base, that may allow pathogens to enter. Here, we propose an alternative model of the relationship between sanitary inspection and microbial water quality, based on the assumption that contamination events occur if three conditions are met: (1) a source of contamination is present, (2) a carrier facilitates the transport of contaminants to the system, and (3) there is a failure in one or more barriers designed to prevent contamination.

Each sanitary risk factor was categorized as a "source", "carrier" or "barrier breakdown" (Fig. 3). Other studies have differentiated sanitary risk factors similarly, including Aldana<sup>33</sup>, who divided sanitary risk factors into "potential sources of fecal contamination; potential routes by which polluting agents gain access to water sources; and factors that can accelerate contamination."

To represent this framework in a statistical model, three binary variables were created to represent the presence of a contamination "source", "carrier", and "barrier breakdown", respectively. For each water system, the binary variable was considered present if any of the sanitary risk factors from the category was present. The binary variable was considered absent if none of the sanitary risk factors in that category were present. For example, the binary variable "source" was considered present if a latrine within 10 meters (m) was present, an uphill latrine was present, another pollution source was present within 10 m of the water source, or any combination of the three; "source" was only considered absent when none of those sanitary risk factors was reported. Logistic regression was used to assess the association between *E. coli* occurrence and every possible combination of "source", "carrier", and "barrier breakdown" variables. We hypothesized that only the scenario where all three variables were present would be associated with *E. coli* occurrence because there would be a break in the causal pathway of contamination if any was absent. In one version of the model, recent rainfall was added to the carrier category using the response to a survey question: "Has it rained in the past two days?" because rainfall is a well-recognized vector of contamination<sup>15,20,34,35</sup>.



**Fig. 3 Theoretical framework to assess sanitary inspection.** Each risk factor included in the WHO sanitary inspection for handpump has been classified as a contamination source, carrier, or breakdown barrier. The framework assumes that at least one contamination source, carrier, and breakdown barrier must all be present to cause water point contamination.

## Model comparison

Models were assessed using Akaike's Information Criterion (AIC)<sup>36</sup>, which assigns a score to each model based on the estimated divergence from the true model. AIC enables a fair comparison among various models and is useful because model simplicity is rewarded. AIC score is a relative model fit measure where a lower AIC value indicates a better model fit.

## DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to them containing information that could compromise research participant privacy.

Received: 27 November 2019; Accepted: 3 November 2020;

Published online: 05 January 2021

## REFERENCES

- WHO/UNICEF. *Progress on Drinking Water, Sanitation and Hygiene*. <http://www.wipo.int/amc/en/> (WHO, 2017).
- Bain, R. et al. Global assessment of exposure to faecal contamination through drinking water based on a systematic review. *Trop. Med. Int. Health* **19**, 917–927 (2014).
- Kumpel, E., Peletz, R., Bonham, M. & Khush, R. Assessing drinking water quality and water safety management in Sub-Saharan Africa using regulated monitoring data. *Environ. Sci. Technol.* **50**, 10869–10876 (2016).
- Pedley, S. & Howard, G. The public health implications of microbiological contamination of groundwater. *Q. J. Eng. Geol. Hydrogeol.* **30**, 179–188 (1997).
- Bain, R. E. S., Wright, J. A., Christenson, E. & Bartram, J. K. Rural: urban inequalities in post 2015 targets and indicators for drinking-water. *Sci. Total Environ.* **490**, 509–513 (2014).
- WHO. *Guidelines for Drinking-Water Quality* 3rd edn (WHO, 2004).
- Cronin, A. A., Breslin, N., Gibson, J. & Pedley, S. Monitoring source and domestic water quality in parallel with sanitary risk identification in Northern Mozambique to prioritise protection interventions. *J. Water Health* **4**, 333–345 (2006).
- WHO. *Guidelines for Drinking-Water Quality* 2nd edn. [http://www.who.int/water\\_sanitation\\_health/dwq/gdwqvol32ed.pdf](http://www.who.int/water_sanitation_health/dwq/gdwqvol32ed.pdf) (WHO, 1997).
- Lloyd, B. J. & Bartram, J. K. Surveillance solutions to microbiological problems in water quality control in developing countries. *Water Sci. Technol.* **24**, 61–75 (1991).
- Okotto-Okotto, J., Okotto, L., Price, H., Pedley, S. & Wright, J. A longitudinal study of long-term change in contamination hazards and shallow well quality in two neighbourhoods of Kisumu, Kenya. *Int. J. Environ. Res. Public Health* **12**, 4275–4291 (2015).
- Luby, S. P. et al. Tubewell water quality and predictors of contamination in three flood-prone areas in Bangladesh. *J. Appl. Microbiol.* **105**, 1002–1008 (2008).
- Lloyd, B. & Suyati, S. A pilot rural water surveillance project in Indonesia. *Waterlines* **7**, 10–13 (1989).
- Misati, A. G., Ogendi, G., Peletz, R., Khush, R. & Kumpel, E. Can sanitary surveys replace water quality testing? Evidence from Kisii, Kenya. *Int. J. Environ. Res. Public Health* **14**, 152 (2017).
- Kelly, E. R., Cronk, R., Kumpel, E., Howard, G. & Bartram, J. How we assess water safety: A critical review of sanitary inspection and water quality analysis. *Sci. Total Environ.* **718**, 137237 (2020).
- Engström, E. et al. Prevalence of microbiological contaminants in groundwater sources and risk factor assessment in Juba, South Sudan. *Sci. Total Environ.* **515–516**, 181–187 (2015).
- Barthiban, S., Lloyd, B. J. & Maier, M. Sanitary hazards and microbial quality of open dug wells in the Maldives Islands. *J. Water Resour. Prot.* **4**, 474–486 (2012).
- Bacci, F. & Chapman, D. V. Microbiological assessment of private drinking water supplies in Co. Cork, Ireland. *J. Water Health* **9**, 738–751 (2011).
- Gerges, D. I., Laplant, W. G., Hyde, J. N., Previl, H. & Forrester, J. Semi-quantitative estimation of *Escherichia coli* levels in public drinking water sources in northern Haiti. *J. Water Sanit. Hyg. Dev.* **6**, 89–95 (2016).
- Tillett, H. E. Potential inaccuracy of microbiological counts from routine water samples. *Water Sci. Technol.* **27**, 15–18 (1993).
- Stukel, T., Greenberg, E., Dain, B., Reed, F. & Jacobs, N. A longitudinal study of rainfall and coliform contamination in small community drinking water supplies. *Environ. Sci. Technol.* **24**, 571–575 (1990).
- Ercumen, A. et al. Can sanitary inspection surveys predict risk of microbiological contamination of groundwater sources? Evidence from shallow tubewells in rural Bangladesh. *Am. J. Trop. Med. Hyg.* **96**, 561–568 (2017).
- Howard, G., Pedley, S., Barrett, M., Nalubega, M. & Johal, K. Risk factors contributing to microbiological contamination of shallow groundwater in Kampala, Uganda. *Water Res.* **37**, 3421–3429 (2003).
- WHO. *Guidelines for Drinking-Water Quality* 1st edn (WHO, 1984).
- Peletz, R., Kumpel, E., Bonham, M., Rahman, Z. & Khush, R. To what extent is drinking water tested in sub-saharan Africa? A comparative analysis of regulated water quality monitoring. *Int. J. Environ. Res. Public Health* **13**, 275 (2016).
- Onda, K., Crocker, J., Kayser, G. L. & Bartram, J. Country clustering applied to the water and sanitation sector: A new tool with potential applications in research and policy. *Int. J. Hyg. Environ. Health* **217**, 379–385 (2014).
- Patrick, J. M., Murcott, S. & Punsalan, J. Coupling microbiological testing and sanitary surveys in drinking water quality programs: results from. *J. Water Sanit. Hyg. Dev.* **01**, 124–135 (2011).
- Snoad, C., Nagel, C., Bhattacharya, A. & Thomas, E. The effectiveness of sanitary inspections as a risk assessment tool for thermotolerant coliform bacteria contamination of rural drinking water: a review of data from West Bengal, India. *Am. J. Trop. Med. Hyg.* **96**, 976–983 (2017).
- Stauber, C., Miller, C., Cantrell, B. & Kroell, K. Evaluation of the compartment bag test for the detection of *Escherichia coli* in water. *J. Microbiol. Methods* **99**, 66–70 (2014).
- WHO. *Guidelines for Drinking-water Quality* 4th edn. <http://apps.who.int/iris/bitstream/handle/10665/254637/9789241549950-eng.pdf;jsessionid=D95B2B826B7C6353F8E8B024DD61400?sequence=1> (WHO, 2017).
- Godfrey, S., Timo, F. & Smith, M. Microbiological risk assessment and management of shallow groundwater sources in Lichinga, Mozambique. *Water Environ. J.* **20**, 194–202 (2006).
- Dey, N. C. et al. Microbial contamination of drinking water from risky tubewells situated in different hydrological regions of Bangladesh. *Int. J. Hyg. Environ. Health* **220**, 621–636 (2017).
- Akoachere, J.-F. T. K., Omam, L.-A. & Massalla, T. N. Assessment of the relationship between bacteriological quality of dug-wells, hygiene behaviour and well characteristics in two cholera endemic localities in Douala, Cameroon. *BMC Public Health* **13**, 692 (2013).
- Aldana, J. *Rapid Assessment of Drinking-water Quality in The Republic of Nicaragua: Country Report of the Pilot Project Implementation in 2004–2005* (WHO, 2010).
- Godfrey, S., Timo, F. & Smith, M. Relationship between rainfall and microbiological contamination of shallow groundwater in Northern Mozambique. *Water SA* **31**, 609–614 (2005).
- Curriero, F. C., Patz, J. A., Rose, J. B. & Lele, S. The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. *Am. J. Public Health* **91**, 1194–1199 (2001).
- Akaike, H. in *Proceedings of the Second International Symposium on Information Theory* (eds Petrov, B. & Caski, F.) 267–281 (1973).

## ACKNOWLEDGEMENTS

We would like to thank World Vision (Federal Way, WA) for financial support. We also thank World Vision Ethiopia, Ghana, Honduras, India, Kenya, Malawi, Mali, Mozambique, Niger, Rwanda, Tanzania, Uganda, Zambia, and Zimbabwe country offices for their support during data collection. Special thanks are due also to the following groups from each of the three countries: World Vision national and area development program office staff, local government officials, community leaders, and committee and community members for sharing their time and experiences with UNC researchers. This study would not have been possible without their participation.

## AUTHOR CONTRIBUTIONS

E.K.: contributed to the conception, data analysis, writing, and revision for this manuscript. R.C.: contributed to the conception of the study, assisted with statistical data analysis, and revision. M.F.: contributed to study design and data collection, writing, and revision. J.B.: contributed substantively to study conception and revision. All authors approve of the submitted manuscript, accept accountability for their contributions, and promise to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41545-020-00093-z>.

**Correspondence** and requests for materials should be addressed to E.K.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021