Water Research 134 (2018) 115–125

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

Efficacy of microbial sampling recommendations and practices in sub-Saharan Africa

David D.J. Taylor [a], Ranjiv Khush [b], Rachel Peletz [c], Emily Kumpel [c, d, *]

[a] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA USA
[b] Aquaya Institute, P.O. Box 5502, Santa Cruz, CA 95063, USA
[c] Aquaya Institute, Riara Corporate Suites, Suite #203, Riara Road, Kilimani Estate, Nairobi, Kenya
[d] Department of Civil and Environmental Engineering, University of Massachusetts Amherst, MA 01003 USA

A R T I C L E I N F O

Article history:
Received 4 October 2017
Received in revised form
18 January 2018
Accepted 23 January 2018
Available online 3 February 2018

Keywords:
Sampling programs
Microbial water quality
Guidelines for drinking water quality
Sub-saharan Africa
Statistical uncertainty
Water quality regulations

A B S T R A C T

Current guidelines for testing drinking water quality recommend that the sampling rate, which is the number of samples tested for fecal indicator bacteria (FIB) per year, increases as the population served by the drinking water system increases. However, in low-resource settings, prevalence of contamination tends to be higher, potentially requiring higher sampling rates and different statistical methods not addressed by current sampling recommendations. We analyzed 27,930 tests for FIB collected from 351 piped water systems in eight countries in sub-Saharan Africa to assess current sampling rates, observed contamination prevalences, and the ability of monitoring agencies to complete two common objectives of sampling programs: determine regulatory compliance and detect a change over time. Although FIB were never detected in samples from 75% of piped water systems, only 14% were sampled often enough to conclude with 90% confidence that the true contamination prevalence met an example guideline ($\leq$5% chance of any sample positive for FIB). Similarly, after observing a ten percentage point increase in contaminated samples, 43% of PWS would still require more than a year before their monitoring agency could be confident that contamination had actually increased. We conclude that current sampling practices in these settings may provide insufficient information because they collect too few samples. We also conclude that current guidelines could be improved by specifying how to increase sampling after contamination has been detected. Our results suggest that future recommendations should explicitly consider the regulatory limit and desired confidence in results, and adapt when FIB is detected.

1. Introduction

While more than 2.6 billion people have gained access to an improved water source over the last 25 years, recent evidence suggests that many of these improved sources do not provide drinking water that is safe (Onda et al., 2012; Bain et al., 2014a; Shaheed et al., 2014; WHO and UNICEF, 2015). Measurements of water quality are important for managing and controlling water safety and tracking progress to national and global targets such as the Sustainable Development Goals (WHO and UNICEF, 2014, 2017). Water management agencies around the world sample microbial drinking water quality to assess whether systems provide water that minimizes risks to health (Rahman et al., 2011; Peletz et al., 2016). Water quality can be monitored for regulatory or operational purposes: regulatory (or verification) monitoring is performed to ensure that a water supply meets standards, while operational monitoring is used to assess operations and detect changes in performance (WHO, 2011). However, collecting samples and testing water quality can be expensive, time-consuming, and logistically complicated (Crocker and Bartram, 2014; Wright et al., 2014; Bain et al., 2014b; Peletz et al., 2016).

Many countries in sub-Saharan Africa (SSA) have adopted or adapted the recommendations in the World Health Organization (WHO) Guidelines for Drinking Water Quality (GDWQ) for the design of their sampling program (Peletz et al., 2016). In practice, many water management agencies in SSA conduct some testing but have not tested enough samples to meet the GDWQ recommendations for the number of samples tested (Peletz et al., 2016). Therefore, optimizing testing is a priority, particularly in low-

* Corresponding author. University of Massachusetts, Amherst, MA, USA. Tel.: +1 413 545 5390.
E-mail address: ekumpel@umass.edu (E. Kumpel).

https://doi.org/10.1016/j.watres.2018.01.054

resource settings where testing activities are constrained and the prevalence of contamination tends to be higher. Previous research on improving the efficacy of sampling has focused on piped supplies in high-income countries that have generous information about their network (e.g. historical data, pipe maps), reliably continuous water supply, and infrequent contamination (Speight et al., 2004; Grayman et al., 2007; van Lieverloo et al., 2007; Horowitz, 2013; Rosen et al., 2009). However, corresponding analyses on improving the effectiveness of sampling in middle- and low-income countries, which often have limited system data, unreliable supplies, and frequent contamination, have not been performed (Lee and Schwab, 2005). To account for the higher prevalence of contamination, more frequent sampling and different statistical methods may be necessary; current sampling recommendations do not address this possibility.

Sampling plans should achieve the goals of a monitoring program while balancing accuracy of results with ease of application and understanding. The GDWQ recommend a minimum number of samples for FIB be tested annually for regulatory monitoring (WHO, 2011); the 2011 GDWQ substantially increased sampling in very large systems (where budgets and health risks may both be larger) and switched from monthly to yearly targets (which may decrease operational costs) as compared to previous editions of the GDWQ (WHO, 1993, 2008). The GDWQ also recommend that none of the tested samples should be found positive for FIB (WHO, 2011). Standards such as these which focus on the allowable number of positive samples are easy to implement and understand, but their statistical power cannot be evaluated. However, in practice, standards that allow no positive samples are indistinguishable from high percentile standards (e.g. $\leq 1\%$ chance of detecting FIB in any sample), whose statistical power we can assess (Ellis, 1989). Similarly, Hunter (2002) compared assessing compliance of bathing water with regulations using a threshold approach (95% of samples complying with water quality standards) to using a percentile approach (the 95th percentile of observed values should not exceed the water quality standard) and found that the percentile method complicated the calculation without changing the regulatory decision. While sampling programs should consider resource and logistical constraints, it is also important to have confidence in the accuracy of the information obtained from a water quality monitoring program, as false positives could lead to expending resources unnecessarily, and false negatives could result in detrimental health consequences.

We analyze what can be learned from past monitoring programs that tested piped water quality in SSA by evaluating each system's: 1) sampling rate and contamination prevalence; 2) ability to assess compliance against a regulatory limit; and 3) ability to detect changes in water quality over time. Our aim is to identify how to increase the effectiveness of sampling recommendations for piped water systems in low-resource settings; we use these results to inform our recommendations for improving current sampling practices and guidelines.

2. Methods

2.1. Data collection

Water quality data were collected as part of the Monitoring for Safe Water (MfSW) program from eight countries in SSA: Benin, Ethiopia, Ghana, Guinea, Kenya, Senegal, Uganda, and Zambia (Peletz et al., 2013, 2016; Kumpel et al., 2016). A full description of the MfSW program and participating agencies are described in Peletz et al. (2016) and Kumpel et al. (2016). Sampling and testing of drinking water was conducted by monitoring agencies responsible for regulatory monitoring of water quality. These included water suppliers (*Supplier*), responsible for providing and monitoring piped water, and health or water surveillance agencies (*Surveillance*), responsible for monitoring and ensuring the quality of all water sources in their jurisdiction. Some information about sampling locations were available for 71% of samples in the database, although these include varying levels of detail (Fig. S1).

Water quality data were collected during two stages of MfSW: 1) retrospective data, collected from agencies that had applied to MfSW (with samples tested between Jan 2009–Dec 2013) and 2) MfSW-supported data, collected by participating monitoring agencies every month (with samples tested between Jul 2013–Apr 2015). The average number of samples taken over time (*sampling rate*) by a given agency increased during the MfSW program since the program provided financial support and incentives to agencies to reach sampling rate targets (Peletz et al., 2013); therefore, the retrospective data represents baseline conditions while the MfSW-supported data represents a 'best-case-scenario' for sampling rate.

2.2. Statistical model for water quality in a piped water system

The microbial quality of piped water varies spatially and temporally throughout a system (Ellis, 1989; Geldreich, 1996). Spatial variations may be introduced by increasing water age, consumption of free chlorine, or point sources of contamination (e.g. backflow or intrusion). Similarly, temporal variations may be induced by changes in source water quality, treatment efficacy, system parameters (e.g. flow rates or pressures), momentary low pressure events (e.g. from maintenance or pressure transients), or time varying contaminant sources (e.g. increased intrusion after rainfall). Any water quality sample that tests positive for FIB provides the utility with specific and conclusive evidence of a problem that existed at a specific time and a specific location in their system. However, from a regulatory perspective, the important question is not the quality of a specific sample at location X and time Y, but, on average, the overall safety of the water distributed by this piped water system (Ellis, 1989). Ellis (1989) suggests four possible ways to account for the temporal and spatial variability of water quality: first, by randomizing samples over space and time; second, by selecting sampling locations representative of 'average' conditions; third, by selecting sampling locations representative of the 'worst' conditions; and finally, by modeling the sources and distribution of contamination and adjusting the sampling strategy accordingly.

Volume 3 of the GDWQ for community supplies recommends that sampling locations should be "representative of the water source, treatment plant, storage facilities, distribution network, points at which water is delivered to the consumer, and points of use" (WHO, 1997). More conservatively, the 2011 GDWQ recommend that samples should be taken at locations with the "best possible chance of detecting contamination" (WHO, 2011). In either case, however, repeated samples from fixed locations can bias the results (Cotter, 1985), and dramatically so where a sampling agency takes action to address contamination after it is detected.

Unfortunately, it was not possible to verify how sampling agencies selected sampling locations. Further, as sampling locations were not uniformly repeated or randomized, credibly accounting for the spatial variations in the samples was beyond the scope of this work. The primary drivers of temporal variations are seasonally- and weather-dependent, both of which are geographically dependent. Given the geographic dispersion in our dataset, accounting for climatic and weather phenomena was beyond the scope of this work.

Therefore, as a starting point for more refined statistical models in the future, and following with analyses conducted in high-income contexts (Ellis, 1989; Cotter, 1985), we combine temporal and spatial sources of variance. Specifically, we model sampling a