ARTICLE OPEN

Exposure measurement error and the characterization of child exposure to fecal contamination in drinking water

Frederick G. B. Goddard^{1✉}, Howard H. Chang², Thomas F. Clasen¹ and Jeremy A. Sarnat¹

Characterizing fecal contamination exposure from drinking water can introduce exposure measurement errors, i.e., differences between the observed and true exposure. These errors can mask the true relationship between fecal contamination exposure and waterborne diseases. We present a framework to quantify the impact of measurement errors on exposure–outcome health effect estimates introduced by variability in measured drinking water fecal contamination levels and household versus community sampling strategies. We matched fecal indicator bacteria (FIB) data for >37,000 drinking water samples to children aged 0–72 months from 19 studies in low- and middle-income countries and took two complementary analytical approaches. We found that household-level exposure assessments may attenuate effect estimates of FIB concentrations in drinking water on diarrhea, and single water samples may attenuate health effect estimates of FIB concentrations on linear growth. To understand the health effects of fecal contamination exposure, measurement error frameworks can be used to estimate more biologically relevant exposures.

npj Clean Water (2020)3:19; <https://doi.org/10.1038/s41545-020-0063-9>

INTRODUCTION

Methods to characterize exposure to fecal contamination from contaminated drinking water in low- and middle-income settings have typically been limited to estimating fecal loading in the environment using indicators of fecal contamination¹. These are proxy measures, in the sense that rather than measuring the actual ingestion of enteric pathogens associated with fecal contamination, they infer exposure by measuring fecal indicator bacteria (FIB) concentrations in the environment and are heavily based on assumptions on the interactions of individuals with that environment. In addition to the well-documented shortcomings of using FIB as a proxy for enteric pathogens², the difference between the observed exposure, i.e., the exposure assigned from these proxy measures, and the true exposure represents a potential form of exposure measurement error. Other areas of environmental health have detailed how exposure measurement error may introduce bias and uncertainty in estimated exposure–outcome relationships, thereby obscuring true associations³. Correspondingly, fields such as air pollution epidemiology have seen increased emphasis on understanding errors associated with differences between individual or personal exposures and other proxy exposure measures used in health effects modeling⁴.

We recently compiled individual participant data (IPD) for a systematic review and meta-analysis. We matched household-level drinking water FIB concentrations as a proxy measure for individual-level enteric pathogen exposure to >43,000 diarrhea reports and >10,000 growth measures for children under the age of 5 years across 19 studies in low- and middle-income countries (Goddard et al., under review). This analysis was an update and expansion of previous meta-analyses that found mixed results on the relationship between FIB concentrations in drinking water and diarrhea and did not consider a more chronic health outcome such as child growth^{5–7}. Our findings suggest that FIB concentrations in household drinking water are associated with both reported diarrhea (odds ratio (OR) 1.09; 95% confidence interval

(CI) 1.04, 1.13) and lower height-for-age Z (HAZ) scores (HAZ –0.04; 95% CI –0.06, –0.01). Notably, we also observed moderate heterogeneity among studies in the strengths of association for both the diarrhea ($I^2 = 34\%$; 95% CI 0–62%) and linear growth analyses ($I^2 = 19\%$; 95% CI 0–63%).

A primary limitation of this analysis was potential error in outcome measurement. In these studies, most of the data are from non-blinded intervention protocols where caregiver-reported diarrhea and linear growth measures are subject to participant and enumerator bias⁸. Another possible source of bias and uncertainty in our analysis may be in part due to errors in the assigned exposure. Prior research studying the effects of exposure measurement error from proxy measures of exposure in air pollution epidemiology suggest that these proxy measures can introduce uncertainty and bias risk estimates toward no observed effect^{9–11}. Similar effects of exposure measurement error have also been shown in other areas, such as chemical exposures¹² and diastolic blood pressure measurements¹³. Based on these findings, it is possible that findings from our IPD analysis may also exhibit similar uncertainty and bias due to exposure measurement error from the use of FIB concentrations in household drinking water as a proxy for personal exposure to enteric pathogens.

Theoretically, there are a number of different potential sources of error in enteric pathogen exposure assessments. Some examples include (1) temporal or spatial variability in water quality, (2) assigning household- or community-level water quality to individuals, (3) the use of FIBs as proxies for enteric pathogens, and (4) processing errors (i.e., during sample collection, transport, or laboratory instrumentation errors). This is not an exclusive list of sources of error, and each source could be further broken down into underlying sources of error. Here we introduce a conceptualized model of potential sources of measurement error using a formal measurement error framework¹⁴. We demonstrate how such a framework might be evaluated and its ability to quantify the relative contributions of measurement error, using empirical

¹Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA. ²Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA. ✉email: frederick.goddard@emory.edu

data of drinking water across several global low- and middle-income settings.

Although quite limited, prior research on the effects of exposure measurement error on waterborne disease epidemiology has found preliminary evidence of regression dilution bias between 14% and 57% from the use of FIB on the relationship between fecal contamination in recreational water and swimming-associated illness¹⁵. Another study found that spatiotemporal variability in rainfall data attenuated subsequent associations between heavy rainfall and diarrhea between 35% and 45%¹⁶. These studies focused on single components of error, and we did not find any formal discussions pertaining to multiple sources of measurement error for exposure to fecal contamination in the peer-reviewed literature. Here we introduce an exposure measurement error framework to conceptualize multiple possible components of error. We adapt a framework for fecal contamination exposure in drinking water based on an approach presented in Zeger et al. to distinguish sources of exposure measurement error¹⁴. While our analysis borrows from extensive work examining air pollution health effects and is limited to fecal contamination as a proxy for enteric pathogens, we intend that this paper serves as an initial step for discussing ways of incorporating estimates of exposure measurement error for enteric pathogen exposures.

Zeger et al. consider sources of error in the assignment of ambient air quality to a population from central monitoring sites, a more distal measure of exposure than attempting to quantify personal exposures. While this framework is contextual to time series studies of air pollution health effects, we see parallels in the proxy measures commonly used for fecal contamination exposure assessments to central-site air pollution exposure assessments, although we do not claim this framework includes all possible sources of error. In the current example involving fecal contamination exposure and response, as in the air pollution design settings that Zeger et al. used, technical and logistical constraints as well as limited resources lead to an inability to obtain measures of true personal exposure x for individual i at time t . Instead, exposure may be estimated by measuring household FIB concentrations z at time t . In Fig. 1, we summarize the differences between measured FIB concentrations z_t , the only component in this framework that is actively measured, and true personal exposure to fecal contamination x_{it} as a proxy for fecal contamination exposure.

There is a difference between measured FIB concentrations in household drinking water z_t and true personal exposure to fecal contamination x_{it} —the exposure measurement error—which we split into three components of error in accordance with the Zeger et al. framework (Eq. 1).

$$x_{it} = z_t + (x_{it} - \bar{x}_t) + (\bar{x}_t - z_t^*) + (z_t^* - z_t) \quad (1)$$

where $(x_{it} - \bar{x}_t)$ describes error from the difference in aggregate fecal contamination exposure across a population $\bar{x}_t$ (i.e., members

of a household) and personal exposure x_{it} ; $(\bar{x}_t - z_t^*)$ describes measurement error from assigning household water fecal contamination z_t^* as the exposure and not considering other exposures, such as exposure to fecal contamination experienced in the community w_t^* that may make up the aggregate exposure $\bar{x}_t$ across a population; and $(z_t^* - z_t)$ describes measurement error from the difference in measured household water FIB concentrations z_t as an indicator of fecal contamination and the true levels of fecal contamination in household drinking water z_t^* .

Our current analysis sought to address the second and third components of this framework with the goal of examining how they may affect exposure–outcome relationships for exposure to fecal contamination in drinking water. This analysis does not seek to validate any single exposure characterization method but rather to describe potential sources of error in current methods to help inform future method development. In conducting this analysis, we sought to assess evidence and magnitude of exposure measurement error from: (1) from assigning household-level FIB concentrations z_t^* as the exposure and not considering community-level FIB concentrations w_t^* (component 2) on health effect estimates of FIB concentrations in drinking water on child diarrhea and (2) from using single FIB measure z_t compared to repeated longitudinal measures z_t^* (component 3) on health effect estimates of FIB concentrations in drinking water on linear growth.

RESULTS

The dataset we compiled included studies of varying sizes with FIB data for drinking water available from 98 to 2137 households per study (Table 1). Four studies included only cross-sectional water sample collection, but most had collected repeated water samples over time with samples typically being collected monthly, quarterly, or annually. To evaluate findings from our simulations, matched diarrhea data were available from all included studies and matched growth data were available from seven studies.

Household versus community exposure

The simulations comparing household to community exposure included 37,119 observations (82% rural, 18% urban) from 16 of the included studies, with between 1 and 12 observations per child. Studies by Arnold et al.²⁶, Brown et al.³⁶, and Pickering et al.⁴⁶ were excluded from these simulations because the data were not compatible for estimating median community water FIB concentrations (i.e., specific sample collection dates were not available or only one sample was collected in every community). The simulations found that if children experienced exposure to fecal contamination in drinking water outside of their household, then using household FIB concentrations z_t^* exclusively as the error-prone exposure variable may lead to an attenuation in the

Fig. 1 Fecal contamination in drinking water exposure measurement error framework (adapted from Zeger et al.¹⁴). This framework outlines three components of error to describe the difference between true personal fecal contamination exposure X_{it} and measured fecal indicator bacteria concentrations Z_t .

Table 1. Summary characteristics of the included studies.

Study ID	Country	No. of households sampled	Samples per household	Matched diarrhea data	Matched growth data
Arnold et al. ³²	India	139	1–3	X	X
Benjamin-Chung et al. ³³	Bangladesh	959	1	X	
Boisson et al. ³⁴	DRC	102	1	X	
Boisson et al. ³⁵	India	2100	1–4	X	
Brown et al. ³⁶	Cambodia	177	1–11	X	
Clasen et al. ³⁷	Colombia	98	1–5	X	
Clasen et al. ³⁸	India	2137	1–8	X	X
Davis (in preparation)	Tanzania	1207	1–6	X	
Ercumen et al. ⁴⁰	Bangladesh	907	1–3	X	
Kirby et al. ⁴¹	Rwanda	265	1–2	X	
Kirby et al. ⁴²	Rwanda	1568	1–3	X	
Luby et al. ²⁹	Bangladesh	810	1–7	X	
Patil et al. ⁴³	India	783	1	X	X
Peletz et al. ⁴⁴	Zambia	232	1	X	
Peletz et al. ⁴⁵	Zambia	116	1–12	X	
Pickering et al. ⁴⁶	Bangladesh	1807	1–3	X	X
Pickering et al. ⁴⁷	Kenya	1986	1–2	X	X
Reese et al. ⁴⁸	India	765	1–4	X	X
Sinharoy et al. ⁴⁹	Rwanda	1744	1–2	X	X

observed FIB–diarrhea relationship even at low levels of community exposure (Fig. 2). If the aggregate exposure $\bar{x}_t$ was represented by 90% household and 10% community exposure, we found that the estimated odds of diarrhea for 1 – $\log_{10}$ higher FIB concentrations in drinking water, using household FIB concentrations in drinking water z_t^* as the error-prone exposure variable, were $OR = 1.32$ compared to the assigned odds of diarrhea $OR = 1.50$ ($\alpha = 0.69$). This trend continued as the assumed aggregate exposure $\bar{x}_t$ consisted of increasing levels of community exposure w_t^* . If the aggregate exposure was represented by 100% community exposure, we found that the estimated odds of diarrhea, using household water as the error-prone exposure variable, were $OR = 1.06$ ($\alpha = 0.15$). Our findings were consistent between urban and rural areas. Sensitivity analyses for these findings are provided in Supplementary Tables S1 and S2.

For the evaluation with empirical diarrhea data, we used the same dataset as we did for the simulations, with the exception of using field-reported diarrhea instead of assigning diarrhea cases. In the combined analysis, we found that household water fecal contamination z_t^* assigned exclusively as the error-prone exposure variable attenuated the association between FIB concentrations in drinking water and diarrhea in comparison to a mixture of household and community water fecal contamination up to assigning 20% household and 80% community water FIB concentrations (Fig. 3). However, this attenuation was not as pronounced as suggested in the simulations, with the greatest attenuation factor found to be $\alpha = 0.79$ (60% household, 40% community exposure). The stratified results suggested that in urban areas the attenuation from using household water fecal contamination z_t^* was limited to scenarios up to 70% household and 30% community water quality, and in rural areas, it extended to 10% household and 90% community water fecal contamination.

After stratifying by age, the odds of diarrhea for 1 – $\log_{10}$ higher FIB concentrations in drinking water for children aged 0–23 months were similar to the odds of diarrhea for children aged 24–72 months when household water fecal contamination z_t^* was assigned as the exposure variable. However, when replacing household water fecal contamination with community

water fecal contamination w_t^* in 10% increments, there was a trend of higher odds of diarrhea for children aged 24–72 months but not for children aged 0–23 months for up to 20% household and 80% community water quality (Fig. 4).

Single versus multiple samples

The simulations comparing the effects of defining exposure with a single water sample compared to multiple samples included 24,806 unique children from the 19 included studies that had ≥ 1 matched FIB concentration estimates in drinking water. These simulations indicated that using FIB concentrations from single water samples z_t compared to the median of multiple samples attenuated estimated differences in HAZ scores associated with fecal contamination in drinking water, with similar findings between wet and dry season months (Fig. 5). If the household water fecal contamination z_t^* was represented by the median of two samples, then randomly selecting one of the two samples almost halved the observed difference in HAZ scores associated with FIB concentrations in drinking water from the assigned $HAZ = -0.20$ to $HAZ = -0.11$ ($\alpha = 0.56$). This finding was more pronounced when household water fecal contamination z_t^* was represented by three ($HAZ = -0.10$; $\alpha = 0.52$) or four samples ($HAZ = -0.09$; $\alpha = 0.43$). Sensitivity analyses for these findings are detailed in Supplementary Tables S3 and S4.

For the evaluation of these simulations, we were limited by empirical linear growth data availability, with linear growth data available for 3311, 743, and 233 children with 2, 3, and ≥ 4 matched water samples, respectively. As a result, we were not able to stratify this analysis by season as we did with the simulations, and the baseline effect estimates and corresponding uncertainties around these estimates vary by group because they represent different samples (Fig. 6). The difference in HAZ scores associated with higher FIB concentrations in drinking water was consistently closer to zero (i.e., no effect) when using a single sample z_t compared to the median of multiple samples. Similar to findings from the simulations, using a single sample compared to the median of two samples approximately halved the estimated difference in HAZ scores associated with FIB concentrations in

Fig. 2 Simulated odds of diarrhea for 1 – $\log_{10}$ higher FIB concentrations in drinking water when only household water FIB concentrations Z_t^* was used as the exposure variable for different household–community exposure scenarios and the corresponding attenuation factors. Exposure scenarios begin with 100% of exposure assumed to be within the household and subsequently replacing household exposure with community exposure W_t^* in 10% increments.

Fig. 3 Estimated odds of diarrhea for 1 – $\log_{10}$ higher FIB concentrations in drinking water for different household–community exposure assignments and the corresponding attenuation factors. Attenuation factors are derived from comparing the log odds of diarrhea for a 100% household exposure assignment to the log odds of diarrhea from community assignments.

drinking water ($\alpha = 0.56$), and this was more pronounced for the median of three or four samples ($\alpha = 0.54$; $\alpha = 0.38$).

DISCUSSION

We adapted and introduced a framework to assess measurement error when characterizing child exposure to fecal contamination in drinking water, i.e., the difference between exposure assigned by proxy measures of exposure and the true exposure experienced by an individual. These frameworks can help prioritize current research gaps by identifying areas within fecal exposure assessments that are limited or missing and by quantifying components of error that are most critical to biases in waterborne disease epidemiology. Ideally, generating improved exposure data could

lead to a better understanding of the true associations between fecal contamination along different pathways and child health. This analysis primarily serves as an initial effort to apply an exposure measurement error framework within the field of waterborne disease epidemiology. In so doing, we aspire to understand the presence and magnitude of several sources of measurement error. Our analyses showed how components of error may attenuate estimated exposure–outcome relationships using empirical data from an extensive dataset of studies collected in low- and middle-income settings. Our findings provide indication that the previously reported odds of diarrhea and reduction in HAZ scores associated with fecal contamination in drinking water (Goddard et al., under review) may be prone to

Estimated odds of diarrhea by household/community fecal contamination assignments:

Fig. 4 Estimated odds of diarrhea for 1 – $\log_{10}$ higher FIB concentrations in drinking water for different household–community exposure assignments, stratified by non-ambulatory and ambulatory pre-school-aged children, and the corresponding attenuation factors. Attenuation factors are derived from comparing the log odds of diarrhea for a 100% household exposure assignment to the log odds of diarrhea from community assignments.

Fig. 5 Simulated difference in HAZ scores for 1 – $\log_{10}$ higher FIB concentrations in drinking water and the corresponding attenuation factors. Simulations compare the use of single measures of household water FIB concentrations z_i as the exposure variable for four scenarios where 1, 2, 3, or 4 samples represent household water fecal contamination z_i^* .

regression dilution bias and thus may be underestimating true exposure–outcome relationships.

We introduced three different components of exposure measurement error. The first component may emerge from assigning household water fecal contamination data to individual household members who interact with their environment differently. Substantial heterogeneity of between-child interactions with their domestic environment has been shown in both urban and rural settings for different age groups in the 0–5-year age range^{17,18}. In addition, differential drinking water ingestion rates by age can lead to heterogeneity in the ingested doses of fecal contamination¹⁹, and infants may experience limited exposure to household water from ingestion before weaning²⁰. This may lead to differences in dose–response between members of the same household. To test how the first component of this exposure

measurement error framework can be applied, small controlled panel studies are needed to generate estimates of personal exposure and compare those to household-level estimates⁹.

The second component of error may occur when exposure to fecal contamination in drinking water outside of the household is not incorporated into exposure assessments. A recent study characterizing fecal exposure in Accra, Ghana as part of the SaniPath research program reported widespread fecal contamination in both domestic and public domains²¹. Measurement error from assigning household water fecal contamination as the exposure does not only depend on the presence of fecal contamination in the public domain but also on the study population's interaction with water in that domain. To our knowledge, no published studies have quantified child exposure to contaminated water in different microenvironments in the

Fig. 6 Estimated difference in HAZ scores with 95% confidence intervals for 1 – $\log_{10}$ higher FIB concentrations in drinking water and the corresponding attenuation factors. Estimates compare randomly selecting single measures of household water FIB concentrations z_t as the exposure variable to estimates with 1, 2, 3, or 4 available samples.

domestic and public domains, but time–activity analyses in air pollution studies have long been conducted for exposure assessments²² and have shown that children spend extensive amounts of time outside of their domestic environment²³.

Findings from our diarrhea simulations suggest that, if children are experiencing exposure to fecal contamination in drinking water outside of their households, then using household water FIB concentrations as a proxy for their overall exposure may result in attenuated health effect estimates for FIB concentrations in drinking water and diarrhea. While evaluating these results, we found that this attenuation may be more pronounced in children above the age of 2 years. This suggests that children under the age of 2 years may be experiencing most of their exposure within the confines of their homes, so household-level exposure assessments may be appropriate for this age group. However, for older ambulatory children exposure outside of the home might be more readily considered.

The third component of error may not only emerge from limited precision associated with methods to characterize FIB concentrations^{24,25}, i.e., from variability in water quality measurements due to sampling and laboratory processing methods, but can also stem from temporal differences in water quality. FIB levels in household water can vary on a weekly, daily, and even hourly basis²⁶. For an outcome such as diarrhea that is normally acute, the biologically relevant household water fecal contamination levels might be representative of the fecal contamination levels during the incubation period of enteric pathogens found in water, which depending on the pathogen can vary from a matter of hours to up to a month²⁷. If water samples are collected on the same day as diarrhea disease data, the measured FIB concentrations on that day may not be representative of the biologically relevant fecal contamination in the lead up to a diarrhea episode. These discrepancies could be due to environmental factors, such as short-term weather changes like extreme rainfall events²⁸, or human factors, such as water treatment behavior change in response to a diarrhea episode^{29,30}. For chronic outcomes such as child growth, the biologically relevant household water fecal contamination likely needs to consider longer-term fecal contamination exposure, which may not be adequately represented by single or a few repeat measurements of household water fecal contamination, due to short-term and seasonal variability in fecal contamination in drinking water³¹.

Our simulations suggest that long-term household water fecal contamination may not be adequately represented by a single sample and hence can result in attenuations of the health effects of FIB concentrations in drinking water on child linear growth. These results were consistent with our evaluations using empirical growth data. While our previously reported IPD analysis found a significant association between fecal contamination and linear growth, 70% of the sample population only had a single matched water quality measure available to characterize exposure. The findings from this analysis imply that the reported effect sizes may be attenuated, and fecal exposure assessments may consider characterizing fecal contamination using multiple longitudinal samples to estimate more biologically relevant exposure.

The results from our analyses need to be interpreted with caution. First, due to data availability we were limited to applying this framework to two sources of error. There are many more possible sources of error that we were not able to consider here, such as assigning household-level exposures to individuals and the use of FIB as proxies for enteric pathogens. Second, this analysis was limited to quantifying the effects of measurement error on the magnitudes of health effects and not on the precision of those effect estimates. Uncertainty in health effect estimates introduced by exposure measurement error may obscure associations where they exist, thus increasing the likelihood of false-negative findings. Third, the current framework is limited to drinking water, but there are a number of other important fecal–oral transmission pathways, such as hands, food, soil, fomites, and flies. Findings from our IPD analysis suggest that fecal contamination along select pathways is associated with child diarrhea and growth, so a similar framework could be applied to other pathways to test whether those findings may have suffered from regression dilution bias. Fourth, we did not have access to repeated water samples within the shorter timeframe of pathogen incubation periods for acute gastroenteritis, so were not able to quantify measurement error on health effects from FIB concentrations in drinking water on diarrhea from the use of single samples used to estimate household water fecal contamination.

Our results suggest that exposure measurement error can contribute to attenuated fecal exposure–outcome relationships for outcomes that are typically acute, such as diarrhea, as well as for more chronic outcomes such as linear growth. Fecal exposure assessments in drinking water may consider exposure outside of the household as well as attempting to characterize fecal

contamination with repeat samples to account for variability in water quality. They may leverage measurement error frameworks to design exposure assessments that are more proximal to the true exposure experienced by individuals, which in turn may inform the design of more effective interventions to reduce waterborne disease burdens.

METHODS

Data

We used data from 19 studies conducted in South America, Sub-Saharan Africa, and South and South-East Asia^{29,32–49}. We requested permission from data owners for use of these data for this study. Eligible datasets included variables describing FIB concentrations in household drinking water, child age, and intervention status. We included children aged 0–72 months. Datasets also included unique anonymized identifiers for each community, household, and child. For the diarrhea analysis, we defined community water fecal contamination levels for a given household on a specific day as the median household water FIB concentrations of all other households in its community on the same day. We generated a variable for each city or collection of communities in rural areas describing whether water quality data was collected in a wet or dry season month by using the 30-year average monthly precipitation from the WorldClim dataset⁵⁰ and designating wet season months as those where average precipitation was >60 mm and dry season months if average precipitation was <60 mm⁵¹. For the diarrhea analysis, we used household water fecal contamination data in long format by matching single household water FIB concentration observations z_t to child survey data if they were collected on the same day or up to 7 days before the survey was conducted and then generated different scenarios for the aggregate fecal contamination $\bar{x}_t$ by incorporating the median community water fecal contamination. For the growth analysis, we used household water FIB concentrations data in wide format by matching all available water samples collected over the course of a child's life up to the day anthropometric measurements were taken.

Analytical approach

We used a two-tiered analytical approach to examine evidence and magnitude of random exposure measurement error. First, in a simulated analysis we randomly assigned health outcomes (diarrhea cases and HAZ scores) to each observation with an estimated exposure and then regressed those outcomes on the error-prone exposure variables, represented by the measured proxies of exposure. Exposure was assigned based on household water FIB concentration measurements, so the simulations retained existing correlations between communities, households, and individuals. Evidence and magnitude of exposure measurement error was assessed by estimating the attenuation factor associated with the error-prone exposure variable⁵². Second, we evaluated findings from our simulations by using empirical health outcome data from the same datasets and regressing it on both the estimated exposure and error-prone exposure variables. All analyses were conducted in R version 3.6⁵³.

Household versus community exposure. We simulated the effect of exclusively assigning household water fecal contamination z_t^* for individual exposure, if estimated exposure is actually a combination of both household and community water fecal contamination w_t^* , by:

1. Randomly generating diarrhea cases for each included child with a combination of household- and community-level drinking water fecal contamination as the aggregate drinking water fecal contamination $\bar{x}_t$ experienced by a child, using the Bernoulli distribution where the log odds of diarrhea d_{ijkl} for child i in household j in community k and study l is given by:

$$\text{logit}(d_{ijkl}) = \beta_0 + \beta_1 \text{FIB}_{ij} + \beta_2 \text{Treat}_{ij} + \beta_3 \text{Age}_i + \beta_4 \text{Season}_{ijk} + \mu_{ijkl} + \mu_{ijk} \quad (2)$$

$$\mu_{ijkl} \sim N(0, 0.6); \mu_{ijk} \sim N(0, 0.3)$$

$$p_{ijkl} = \frac{e^{d_{ijkl}}}{1 + e^{d_{ijkl}}}$$

$$d_{ijkl} = \text{Bernoulli}(ijkl, p_{ijkl})$$

2. Assuming that (1) community-level drinking water fecal contamination is represented by the median household water FIB concentrations in all other community households; (2) baseline odds of diarrhea β_0 for this population is 0.15; (3) odds of diarrhea for 1 – $\log_{10}$ higher FIB concentrations in drinking water β_1 is 1.5; and (4) odds of diarrhea for children receiving an intervention β_2 , child age β_3 (in years), and for data collected in the wet compared to the dry season β_4 are 0.9, 0.8, and 1.2, respectively. Effect estimates were broadly based on model outputs from our IPD analyses, although we assumed higher odds of diarrhea for FIB concentrations in drinking water because we hypothesize that the effect estimate for the exposure–outcome relationship in our IPD analysis may have been suffering from regression dilution bias⁵⁴. The model accounted for clustering at the study-level μ_{ijkl} and community-level μ_{ijk} .
3. Fitting multilevel generalized mixed effects models with the assigned diarrheal cases and replacing the combined household and community drinking water FIB concentrations $\bar{x}_t$ with household-level FIB concentrations z_t^* exclusively as the error-prone exposure variable.
4. Calculating the attenuation associated with the estimated log odds of diarrhea (β_1^*) from assigning household-level FIB concentrations z_t^* exclusively as the error-prone exposure variable, compared to the assigned log odds of diarrhea ($\beta_1 = \log(1.5)$) if combined household and community drinking water FIB concentrations $\bar{x}_t$ represent the exposure: $a = \frac{\beta_1^*}{\beta_1} = \frac{\beta_1^*}{\log(1.5)}$.
5. Repeating simulations for a range of exposure scenarios by adding community water fecal contamination w_t^* in 10% increments, starting with 100% household water fecal contamination and ending with 100% community water fecal contamination representing the estimated exposure.
6. Stratifying the combined analysis: As reported previously, estimated odds of diarrhea for a 1 – $\log_{10}$ increase in FIB concentrations in drinking water was higher in urban compared to rural settings (Goddard et al., under review), so we stratified the simulation by urban versus rural areas to differentiate whether exclusively assigning household drinking water fecal contamination z_t^* may introduce more error in one setting compared to the other.
7. Conducting sensitivity analyses: Assessed the effects our assumptions had on the simulation findings by repeating the simulations with (a) higher and lower assumed odds of diarrhea for higher FIB concentrations in drinking water and (b) using the highest and lowest community water FIB concentrations instead of the median.

To evaluate findings from the simulations, we applied empirical diarrhea data by:

1. Beginning with household water fecal contamination z_t^* as the estimated exposure and fitting a multilevel generalized mixed effects model to estimate the odds of diarrhea for 1 – $\log_{10}$ higher FIB concentrations in household drinking water.
2. Replacing household water fecal contamination z_t^* with community water fecal contamination w_t^* in 10% increments and fitting the same regression model with each new exposure assignment.
3. Calculating the attenuation associated with the log odds of diarrhea (β_1^*) from assigning household-level FIB concentrations exclusively as the error-prone exposure variable, compared to effect estimates that combine household/community water fecal contamination (β_1).
4. Stratifying the analysis: In addition to stratifying by rural versus urban areas, we also stratified by children aged 0–23 and 24–72 months to consider how child mobility may modify the effect of assigning community water quality to exposure. We hypothesized that children aged 0–23 months are mostly non-ambulatory and spend the majority of their time within the confines of their home, and pre-school children aged 24–72 months are ambulatory and spend their time both in their home and within the confines of the community.

Single versus multiple samples. We simulated the effect of assigning a single measure of FIB concentrations in drinking water z_t as the error-prone exposure variable by

1. Randomly generating expected HAZ scores with the estimated household water fecal contamination z_t^* represented by the median household water FIB concentrations from repeat samples, using the following model where the difference in HAZ scores HAZ_{ijk} for child

i in household j in community k and study l is given by:

$$\text{HAZ}_{ijkl} = \beta_0 + \beta_1 \text{FIB}_{ij} + \beta_2 \text{Treat}_{ij} + \beta_3 \text{Age}_i + \mu_{ijkl} + \epsilon_{ijkl}; \quad (3)$$

$$\mu_{ijkl} \sim N(0, 0.5); \epsilon_{ijkl} \sim N(0, 1)$$

- Assuming that the (1) mean baseline HAZ score β_0 in this population is -1.6 ; (2) difference in HAZ score for $1 - \log_{10}$ higher median FIB concentrations β_1 is -0.2 ; (3) difference in HAZ score for children receiving an intervention β_2 and for child age β_3 (years) are 0.1 and -0.05 , respectively; and (4) HAZ scores follow a normal distribution. The model accounted for clustering at the study-level μ_{ijkl} .
- Fitting multilevel generalized mixed-effects models with the assigned HAZ scores and replacing the estimated household water fecal contamination z_i^* represented by the median household water FIB concentrations from repeat samples with a randomly chosen single measure of household water FIB concentrations z_i as the error-prone exposure variable.
- Calculating the attenuation associated with the estimated difference in HAZ score (β_1^*) from randomly choosing a single measure of water quality z_i as the error-prone exposure variable, compared to the assigned difference in HAZ score ($\beta_1 = -0.2$), if the estimated exposure is represented by repeat samples of household water fecal contamination z_i^* : $\alpha = \frac{\beta_1^*}{\beta_1} = \frac{\beta_1}{-0.2}$.
- Repeating the simulations for children with at least two, three, or four matched household water FIB concentration measures making up the median household water fecal contamination z_i^* . We did not have sufficient data to conduct these simulations with more than four matched water samples.
- Stratifying the analysis: Previous research has found that fecal contamination in drinking water sources in low-income countries is higher in the wet season compared to the dry season⁵⁵, so we stratified these simulations by season to examine whether error introduced from variability in water quality is greater in one season compared to the other.
- Conducting sensitivity analyses: Assessed the effects our assumptions had on the simulation findings by repeating the simulations with (a) higher and lower assumed difference in HAZ scores for higher FIB concentrations in drinking water and (b) using the highest and lowest drinking water FIB concentrations from the repeat samples instead of the median.

To evaluate findings from the simulations, we applied empirical linear growth data for a subset of children in our dataset where HAZ scores were available by:

- Fitting multilevel generalized mixed-effects models with the median household water fecal contamination z_i^* from repeated measures of household FIB concentrations as the exposure variable. Repeating this for two, three, and four repeat measures.
- Fitting the same models after randomly selecting a single measure of household water FIB concentrations z_i as the error-prone exposure variable from the repeat measures.
- Calculating the attenuation associated with the estimated difference in HAZ score (β_1^*) from randomly selecting a single measure of household water FIB concentrations z_i as the error-prone exposure variable, compared to effect estimate (β_1) from the median of repeat samples of household water fecal contamination z_i^* as the exposure variable.
- Stratifying the analysis: We conducted the same stratification by season as we did for the simulations.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding authors of the included studies, but restrictions apply to the availability of these data. Most of the included data are not publicly available; however, data are available from the authors upon reasonable request. The corresponding author of this study does not own any of the included data.

CODE AVAILABILITY

The code used for the analyses is available from the corresponding author upon request.

Received: 27 September 2019; Accepted: 6 March 2020;

Published online: 06 April 2020

REFERENCES

- Sciar, G. D. et al. Assessing the impact of sanitation on indicators of fecal exposure along principal transmission pathways: a systematic review. *Int. J. Hyg. Environ. Health* **219**, 709–723 (2016).
- Leclerc, H., Mossel, D. A. A., Edberg, S. C. & Struijk, C. B. Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. *Annu. Rev. Microbiol.* **55**, 201–234 (2001).
- Armstrong, B. G. Effect of measurement error on epidemiological studies of environmental and occupational exposures. *Occup. Environ. Med.* **55**, 651–656 (1998).
- Sarnat, J. A. et al. Panel discussion review: session 1—exposure assessment and related errors in air pollution epidemiologic studies. *J. Expo. Sci. Environ. Epidemiol.* **17**, S75–S82 (2007).
- Gundry, S., Wright, J. & Conroy, R. A systematic review of the health outcomes related to household water quality in developing countries. *J. Water Health* **2**, 1–13 (2004).
- Gruber, J. S., Ercumen, A. & Colford, J. M. Coliform bacteria as indicators of diarrheal risk in household drinking water: systematic review and meta-analysis. *PLoS ONE* **9**, e107429 (2014).
- Hodge, J. et al. Assessing the association between thermotolerant coliforms in drinking water and diarrhea: an analysis of individual-level data from multiple studies. *Environ. Health Perspect.* **124**, 1560–1567 (2016).
- Schmidt, W.-P. et al. Epidemiological methods in diarrhoea studies—an update. *Int. J. Epidemiol.* **40**, 1678–1692 (2011).
- Schwartz, J., Sarnat, J. A., Coull, B. A. & Wilson, W. E. Effects of exposure measurement error on particle matter epidemiology: a simulation using data from a panel study in Baltimore, MD. *J. Expo. Sci. Environ. Epidemiol.* **17**, S2–S10 (2007).
- Goldman, G. T. et al. Impact of exposure measurement error in air pollution epidemiology: effect of error type in time-series studies. *Environ. Health* **10**, 61 (2011).
- Pennington, A. F. et al. Measurement error in mobile source air pollution exposure estimates due to residential mobility during pregnancy. *J. Expo. Sci. Environ. Epidemiol.* **27**, 513–520 (2017).
- Perrier, F., Giorgis-Allemand, L., Slama, R. & Philippat, C. Within-subject pooling of biological samples to reduce exposure misclassification in biomarker-based studies. *Epidemiology* **27**, 378–88 (2016).
- MacMahon, S. et al. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* **335**, 765–74 (1990).
- Zeger, S. L. et al. Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ. Health Perspect.* **108**, 419–426 (2000).
- Fleisher, J. M. The effects of measurement error on previously reported mathematical relationships between indicator organism density and swimming-associated illness: a quantitative estimate of the resulting bias. *Int. J. Epidemiol.* **19**, 1100–1106 (1990).
- Levy, M. C. et al. Spatiotemporal error in rainfall data: consequences for epidemiologic analysis of waterborne diseases. *Am. J. Epidemiol.* **188**, 950–959 (2019).
- Teunis, P. F. M., Reese, H. E., Null, C., Yakubu, H. & Moe, C. L. Quantifying contact with the environment: behaviors of young children in Accra, Ghana. *Am. J. Trop. Med. Hyg.* **94**, 920–931 (2016).
- Kwong, L. H. et al. Age-related changes to environmental exposure: variation in the frequency that young children place hands and objects in their mouths. *J. Expo. Sci. Environ. Epidemiol.* <https://doi.org/10.1038/s41370-019-0115-8> (2019).
- US EPA. *Exposure Factors Handbook: 2011 Edition* (EPA, 2011).
- VanDerslice, J., Popkin, B. & Briscoe, J. Drinking-water quality, sanitation, and breast-feeding: their interactive effects on infant health. *Bull. World Health Organ.* **72**, 589–601 (1994).
- Robb, K. et al. Assessment of fecal exposure pathways in low-income urban neighborhoods in Accra, Ghana: rationale, design, methods, and key findings of the SaniPath Study. *Am. J. Trop. Med. Hyg.* **97**, 1020–1032 (2017).
- Branco, P. T. B. S., Alvim-Ferraz, M. C. M., Martins, F. G. & Sousa, S. I. V. The microenvironmental modelling approach to assess children's exposure to air pollution – a review. *Environ. Res.* **135**, 317–332 (2014).
- Devakumar, D. et al. Biomass fuel use and the exposure of children to particulate air pollution in southern Nepal. *Environ. Int.* **66**, 79–87 (2014).
- Harmel, R. D. et al. Uncertainty in monitoring *E. coli* concentrations in streams and stormwater runoff. *J. Hydrol.* **534**, 524–533 (2016).
- Gronewold, A. D. & Wolpert, R. L. Modeling the relationship between most probable number (MPN) and colony-forming unit (CFU) estimates of fecal coliform concentration. *Water Res.* **42**, 3327–3334 (2008).

26. Levy, K., Hubbard, A. E., Nelson, K. L. & Eisenberg, J. N. S. Drivers of water quality variability in Northern Coastal Ecuador. *Environ. Sci. Technol.* **43**, 1788–1797 (2009).
27. European Centre for Disease Prevention and Control. *Systematic Review on the Incubation and Infectiousness/Shedding Period of Communicable Diseases in Children* (European Centre for Disease Prevention and Control, 2016).
28. Guzman Herrador, B. R. et al. Analytical studies assessing the association between extreme precipitation or temperature and drinking water-related waterborne infections: a review. *Environ. Health* **14**, 29 (2015).
29. Luby, S. P. et al. Microbiological contamination of drinking water associated with subsequent child diarrhea. *Am. J. Trop. Med. Hyg.* **93**, 904–911 (2015).
30. Ercumen, A. et al. Potential sources of bias in the use of *Escherichia coli* to measure waterborne diarrhoea risk in low-income settings. *Trop. Med. Int. Health* **22**, 2–11 (2016).
31. Wilkes, G. et al. Seasonal relationships among indicator bacteria, pathogenic bacteria, *Cryptosporidium* oocysts, *Giardia* cysts, and hydrological indices for surface waters within an agricultural landscape. *Water Res.* **43**, 2209–2223 (2009).
32. Arnold, B. F. et al. Causal inference methods to study nonrandomized, preexisting development interventions. *Proc. Natl Acad. Sci. USA* **107**, 22605–22610 (2010).
33. Benjamin-Chung, J. et al. A randomized controlled trial to measure spillover effects of a combined water, sanitation, and handwashing intervention in rural Bangladesh. *Am. J. Epidemiol.* **187**, 1733–1744 (2018).
34. Boisson, S. et al. Field assessment of a novel household-based water filtration device: a randomised, placebo-controlled trial in the democratic Republic of Congo. *PLoS ONE* **5**, 1–10 (2010).
35. Boisson, S. et al. Effect of household-based drinking water chlorination on diarrhoea among children under five in Orissa, India: a double-blind randomised placebo-controlled trial. *PLoS Med.* **10**, e1001497 (2013).
36. Brown, J., Sobsey, M. D. & Loomis, D. Local drinking water filters reduce diarrheal disease in Cambodia: a randomized, controlled trial of the ceramic water purifier. *Am. J. Trop. Med. Hyg.* **79**, 394–400 (2008).
37. Clasen, T., Parra, G. G., Boisson, S. & Collin, S. Household-based ceramic water filters for the prevention of diarrhea: a randomized, controlled trial of a pilot program in Colombia. *Am. J. Trop. Med. Hyg.* **73**, 790–795 (2005).
38. Clasen, T. et al. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *Lancet Glob. Health* **2**, e645–e653 (2014).
39. Mattioli, M. C. et al. Enteric pathogens in stored drinking water and on caregiver's hands in Tanzanian households with and without reported cases of child diarrhea. *PLoS ONE* **9**, e84939 (2014).
40. Ercumen, A. et al. Effects of source- versus household contamination of tubewell water on child diarrhea in rural Bangladesh: a randomized controlled trial. *PLoS ONE* **10**, e0121907 (2015).
41. Kirby, M. A. et al. Use, microbiological effectiveness and health impact of a household water filter intervention in rural Rwanda—a matched cohort study. *Int. J. Hyg. Environ. Health* **220**, 1020–1029 (2017).
42. Kirby, M. A. et al. Effects of a large-scale distribution of water filters and natural draft rocket-style cookstoves on diarrhea and acute respiratory infection: a cluster-randomized controlled trial in Western Province, Rwanda. *PLoS Med.* **16**, e1002812 (2019).
43. Patil, S. R. et al. The effect of India's total sanitation campaign on defecation behaviors and child health in rural Madhya Pradesh: a cluster randomized controlled trial. *PLoS Med.* **11**, e1001709 (2015).
44. Peletz, R. et al. Drinking water quality, feeding practices, and diarrhea among children under 2 years of HIV-positive mothers in peri-urban Zambia. *Am. J. Trop. Med. Hyg.* **85**, 318–26 (2011).
45. Peletz, R. et al. Assessing water filtration and safe storage in households with young children of HIV-positive mothers: a randomized, controlled trial in Zambia. *PLoS ONE* **7**, e46548 (2012).
46. Pickering, A. J. et al. Fecal indicator bacteria along multiple environmental transmission pathways (water, hands, food, soil, flies) and subsequent child diarrhea in rural Bangladesh. *Environ. Sci. Technol.* **52**, 7928–7936 (2018).
47. Pickering, A. et al. Can individual and integrated water, sanitation, and hand-washing interventions reduce fecal contamination in the household environment? Evidence from the WASH Benefits cluster-randomized trial in rural Kenya. Preprint at *BiorXiv* <https://doi.org/10.1101/731992> (2019).
48. Reese, H. et al. Assessing longer-term effectiveness of a combined household-level piped water and sanitation intervention on child diarrhoea, acute respiratory infection, soil-transmitted helminth infection and nutritional status: a matched cohort study in rural Odisha. *Int. J. Epidemiol.* **0**, 1–11 (2019).
49. Sinharoy, S. S. et al. Effect of community health clubs on child diarrhoea in western Rwanda: cluster-randomised controlled trial. *Lancet Glob. Health* **5**, e699–e709 (2017).
50. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
51. Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11**, 1633–1644 (2007).
52. Fuller, W. A. *Measurement Error Models* (John Wiley & Sons, 1987).
53. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2019).
54. Hutcheon, J. A., Chiolero, A. & Hanley, J. A. Random measurement error and regression dilution bias. *BMJ* **340**, c2289 (2010).
55. Kostyla, C., Bain, R., Cronk, R. & Bartram, J. Seasonal variation of fecal contamination in drinking water sources in developing countries: a systematic review. *Sci. Total Environ.* **514**, 333–343 (2015).

ACKNOWLEDGEMENTS

We thank Amy Pickering, Ayse Ercumen, Joe Brown, Ben Arnold, Jade Benjamin-Chung, Sophie Boisson, Jenna Davis, Angela Harris, Wolf-Peter Schmidt, Steve Luby, Miles Kirby, Sumeet Patil, Rachel Peletz, Heather Reese, and Sheela Sinharoy for providing data for this study.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception of this study. J.A.S., F.G.B.G., and H.H.C. developed the analysis plan. F.G.B.G. conducted the systematic search and data compilation for the parent study that identified the data used in this study. F.G.B.G. conducted the data analysis and wrote the first draft of the manuscript. T.F.C. contributed original data to this study. All authors contributed to the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41545-020-0063-9>.

Correspondence and requests for materials should be addressed to F.G.B.G.

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) 2020