International Journal of Hygiene and Environmental Health 220 (2017) 29–36

Contents lists available at ScienceDirect

International Journal of Hygiene and Environmental Health

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

# Quantifying tap-to-household water quality deterioration in urban communities in Vellore, India: The impact of spatial assumptions☆

Tania M. Alarcon Falconi[a], Alexandra V. Kulinkina[a], Venkata Raghava Mohan[b], Mark R. Francis[c], Deepthi Kattula[c], Rajiv Sarkar[c], Honorine Ward[c,d], Gagandeep Kang[c], Vinohar Balraj[b], Elena N. Naumova[a,c,e,*]

[a] Department of Civil & Environmental Engineering, Tufts University, Medford, MA, USA
[b] Department of Community Health, Christian Medical College, Vellore, Tamil Nadu, India
[c] Division of Gastrointestinal Sciences, Christian Medical College, Vellore, Tamil Nadu, India
[d] Department of Geographic Medicine, Tufts Medical Center, Boston, MA, USA
[e] Friedman School of Nutrition Science & Policy, Tufts University, Boston, MA, USA

ARTICLE INFO

*Article history:*
Received 29 June 2016
Received in revised form
26 September 2016
Accepted 26 September 2016

*Keywords:*
Water quality
Spatial modeling
Spatial uncertainty
Household water storage
Waterborne diseases
GIS

ABSTRACT

Municipal water sources in India have been found to be highly contaminated, with further water quality deterioration occurring during household storage. Quantifying water quality deterioration requires knowledge about the exact source tap and length of water storage at the household, which is not usually known. This study presents a methodology to link source and household stored water, and explores the effects of spatial assumptions on the association between tap-to-household water quality deterioration and enteric infections in two semi-urban slums of Vellore, India. To determine a possible water source for each household sample, we paired household and tap samples collected on the same day using three spatial approaches implemented in GIS: minimum Euclidean distance; minimum network distance; and inverse network-distance weighted average. Logistic and Poisson regression models were used to determine associations between water quality deterioration and household-level characteristics, and between diarrheal cases and water quality deterioration. On average, 60% of households had higher fecal coliform concentrations in household samples than at source taps. Only the weighted average approach detected a higher risk of water quality deterioration for households that do not purify water and that have animals in the home (RR = 1.50 [1.03, 2.18], p = 0.033); and showed that households with water quality deterioration were more likely to report diarrheal cases (OR = 3.08 [1.21, 8.18], p = 0.02). Studies to assess contamination between source and household are rare due to methodological challenges and high costs associated with collecting paired samples. Our study demonstrated it is possible to derive useful spatial links between samples *post hoc*; and that the pairing approach affects the conclusions related to associations between enteric infections and water quality deterioration.



## 1. Introduction

According to the United Nations, over 90 percent (%) of the world population use improved drinking water sources with over half of the population using water that is piped on premises (WHO/UNICEF, 2014). Although improved water sources are considered to provide some protection against water contamination, 38% of water quality studies in low- and middle-income countries report that over a quarter of the samples have fecal contamination (Bain et al., 2014b). Additionally, over 10% of improved sources may contain fecal bacteria concentrations in excess of 100 colony forming units per 100 milliliters (CFU/100 ml; Bain et al., 2014a; WHO/UNICEF, 2014). Even if source water quality conforms with World Health Organization (WHO) guidelines (not detectable in any 100 ml sample, WHO, 2011), there is a risk of contamination during transport and storage in areas with intermittent water supply (Bhunia et al., 2009; Brick et al., 2004; Wright et al., 2004). Over 300 million people worldwide receive piped water intermit-

☆ The study was approved by the Institutional Review Boards (IRB) at Christian Medical College (CMC), Vellore, India and Tufts University School of Medicine, Boston, USA. All study activities were carried out in accordance with the approved guidelines.

* Corresponding author at: Friedman School of Nutrition Science & Policy, 150 Harrison Ave, Jaharis-269F, Boston, MA, USA.
*E-mail address:* elena.naumova@tufts.edu (E.N. Naumova).

http://dx.doi.org/10.1016/j.ijheh.2016.09.019
1438-4639/

tently, between a few hours a day to a few hours a week, with South Asian developing countries having on average 7 hours of water supply per day (Bhunia et al., 2009; Biswas, 2007; Brick et al., 2004; Kumpel and Nelson, 2016). In India, approximately 19% of households in urban slums have access to piped water, compared to 50% of all urban households. However, over 20% of urban dwellers still obtain their water from a source located 100 meters or more from their premises (UNICEF/FAO/SaciWATERs, 2013). Municipal water sources in India have been found to have high levels of microbial contamination, with further degradation of water quality during household storage (Brick et al., 2004; Firth et al., 2010; Sarkar et al., 2007; Trevett et al., 2005), resulting in a high burden of water-related diseases (Prüss-Ustün et al., 2014; Sarkar et al., 2013a,b).

Numerous studies have reported associations of changes in water contamination levels between the source and household with environmental parameters and household-level behavioral factors (Eshcol et al., 2009; Levy et al., 2008; Mintz et al., 1995). Crowding (people/room), water purification, and latrine use have been associated with an increased risk of water quality deterioration from source to household (Brick et al., 2004; Kattula et al., 2015). Studies of water quality deterioration rely on either paired or separate water collection at the source and household (Shields et al., 2015). Paired sample collection, where water is followed from the source to the storage container, provides detailed information regarding the source of water, storage conditions, and potential sources of stored-water contamination (Levy et al., 2008). However, even with well-designed sample collection methods, intermittent water supply may result in field conditions guiding sampling schemes, with source sampling being possible only during the limited supply periods. Studies that collect source and household water separately provide flexibility on time of water collection (Abdellah et al., 2012; Aldana 2010; Aliev et al., 2010). However, such a study design requires matching of sources and households *post hoc* (Shields et al., 2015) and usually results in limited information regarding the exact water source and length of household water storage. Scientists may nonetheless rely on data from samples collected separately when such data are readily available. The present study utilizes secondary data to explore the effects of multiple water sample pairing methods on the associations among water quality deterioration, diarrheal cases and household characteristics.

The data for the study was obtained from a project funded by the Centers for Disease Control and Prevention and the Indian Council of Medical Research designed to assess the relationship between enteric infections and environmental parameters in Vellore, India (Kattula et al., 2015; Kulinkina et al., 2016). Disease surveillance was conducted concurrently with water quality monitoring from household storage containers and public piped water supply taps, henceforward referred to as taps. Public water supply showed widespread microbial contamination and diarrhea incidence rates were associated with behavioral and environmental parameters such as presence of animals in the household (Kattula et al., 2015; Kulinkina et al., 2016). Although characterizing tap-to-household water quality deterioration was not an objective of the main study, the present analysis was made possible due to the study area being fully geocoded (Sarkar et al., 2007) and the availability of dates associated with all observations. Because water samples were collected from taps and households on separate schedules, it was not possible to know for certain the spatial and temporal links between tap and household samples. We therefore present a methodology that can be used to link source and household samples using multiple spatial approaches and present the effect of the uncertainty associated with each approach on diarrhea risk estimations.

In any study design, the ability to verify and validate the exposure is crucial for establishing causal relationships. In observational studies aiming to examine associations between household water quality and health outcomes, characterizing exposure requires links to be established between source and household water. In this analysis we formed a temporal link (household water was collected on the same day it was sampled), and used three assumptions (source taps selected based on different measures of spatial proximity) to form spatial links. Thus, the overarching goal of this study is to present a methodology that explores the effects of the spatial assumptions on the association between tap-to-household water quality deterioration and enteric infections in two semi-urban slums of Vellore, India. The specific aims of the study are to: (1) characterize tap-to-household water quality deterioration under different spatial assumptions; (2) analyze the association between water quality deterioration and household characteristics; and (3) analyze the association between diarrheal episodes and water quality deterioration. This study allowed us to systematically assess and demonstrate the effects of different measures of proximity on the association of enteric infections and tap-to-household water quality deterioration.

## 2. Methods

### 2.1. Characterization of tap-to-household deterioration

Water quality was monitored between August 2010 and March 2012 in 36 public taps and 160 household water storage containers in two geographically adjacent semi-urban slums of Vellore, India. Study households were recruited using staggered enrollment based on eligibility criteria; and study taps were selected at random (Kulinkina et al., 2016). Frequency of sample collection ranged from once a month for taps to four times per year for households. Water samples were analyzed by the Wellcome Research Laboratory at the Christian Medical College (CMC), Vellore India using standard diagnostic kits and methods. Fecal coliforms in CFU/100 ml were quantified using M-FC media (HiMedia Labs Pvt. Ltd.) in two dilutions (direct and 1:10), and final concentrations were calculated by averaging all dilution plates. If concentrations were too numerous to count, the dilutions were replaced with a value twice the upper limit of the countable range of the highest dilution. Chlorine residual was not detected in any of the water samples. Kulinkina et al. (2016) provide a detailed description of the study sites, sampling procedure, and water quality analysis methods, including sample selection, handling, and sampling seasons.

A source of uncertainty in our study was lack of information on the time when each household collected water from a source tap. To establish temporal tap-to-household links and to minimize the associated error, we first selected households that had stored water sampled on the same day as available tap water samples. Since the study area was fully geocoded, we used a geographic information system (GIS) to manually create road networks. We formed links between all study taps and households and determined two values of proximity between them: Euclidean (straight-line) and network (length of path) distance. If a household had more than one possible source tap in a given day, we relied on three spatial approaches to create a single tap-to-household link. Spatial approach A selected the link with the minimum Euclidean tap-to-household distance; approach B selected the link with the minimum network tap-to-household distance; and approach C estimated source water quality by calculating a weighted average of all possible source taps using inverse network-distance as the weight. Fig. 1 shows a schematic of the three spatial approaches.

Once tap-to-household links were formed, fecal coliform measurements were $\log_{10}$ transformed. Tap-to-household deterioration was calculated as the difference between $\log_{10}$ transformed fecal coliform concentrations at the source tap and at the household. A binary deterioration outcome was defined as 1 for a positive

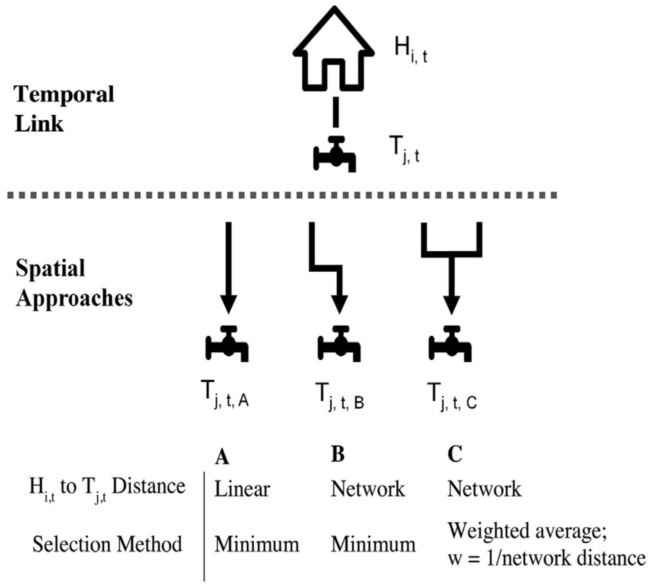


**Fig. 1.** Schematic of three spatial approaches (A, B, and C). Households ($H_i$) and potential source taps ($T_j$) are first selected based on same-day sampling (t), and then source taps are selected based on measures of proximity (Euclidean and network distances).

difference and 0 for negative or zero difference (i.e., contamination at the household water improved or showed no change compared to the source tap). Approximately 40% of all observations showed no deterioration. $Log_{10}$ differences were also rounded to the closest integer and coded as an ordered outcome where level 0 indicates no deterioration and levels 1 through 3 indicate a $log_{10}$ increase in fecal contamination; for example, observations in deterioration level 1 represent households that had fecal coliform concentrations between 0 and 1 $log_{10}$ higher than the source tap. These levels were selected for ease of interpretation. Since some of the differences may be due to laboratory or sampling variation rather than actual water quality deterioration, these levels constitute thresholds of potential water quality deterioration.

Presence or absence of tap-to-household deterioration for each spatial approach was summarized by number of samples and their percentage, and by mean and standard deviation for ordered levels; and the difference between approaches A, B, and C was tested by analysis of variance (ANOVA) for continuous variables and Pearson's chi-squared test for categorical variables.

### 2.2. Association between tap-to-household deterioration and household characteristics

Household demographic and behavioral parameters were collected at enrollment (Kattula et al., 2015) and included information on water purification, water used for washing fruits/vegetables, use of latrines, handwashing with soap, presence of animals in the household, animal shed inside the home, and frequency of children playing with animals. The responses were coded as 0 or 1 to reflect potential low or high risk behavior; for example children playing with animals on most days was coded as 1 (high risk behavior) and not on most days as 0 (low risk behavior). All characteristics were tested for correlation with tap-to-household deterioration for all spatial approaches (A, B, and C) using ANOVA. Based on the lowest *p*-values (Supplemental material, Table S1), two of these characteristics (water purification and animal presence in the household) were aggregated into an "exposure score" ranging from 0 (low risk behavior) to 2 (high risk behavior) and examined in further analysis. A high risk behavior of 2, for example, constituted households that do not purify water and have animals in the home. Additional household characteristics included in this study were crowding (people/room) and household size, either total or based on number of people in each age category. There were six different age categories: young children (<5 years old), children (5 to 15), young adults (16 to 40), adults (41 to 60), and elderly (>60). Because temporal variability in tap and household water quality has been demonstrated in a prior study (Kulinkina et al., 2016), we tested for seasonality in tap-to-household deterioration based on the local Tamil calendar. Only the Chill/Cold Tamil season (October 15 to December 14, ~26 °C average temperature, and ~40 mm average precipitation) was significantly correlated ($p < 0.5$) with tap-to-household deterioration and was examined in further analysis. The detected seasonal variability was accounted in regression models by including a binary variable reflecting the time of sampling (coded as 1, for samples collected during October 15 to December 14, which refers to Chill/Cold Tamil season, and as 0, otherwise). Exposure score, crowding, household size, and time of sampling were included as additional household-level covariates.

To explore the association between fecal coliform increase, as a binary and as an ordered outcome, and household characteristics, we first performed an exploratory data analysis comparing household characteristics in houses with and without deterioration for each spatial approach. Descriptive statistics were stratified by presence or absence of tap-to-household fecal coliform increase and statistical significance between groups was determined by ANOVA for continuous variables and Pearson's chi-squared test for categorical variables.

To further investigate the association between tap-to-household deterioration as assessed by each spatial approach (A, B, and C) and household characteristics, two types of regression models were used. A logistic fixed effects regression model was applied to a binary outcome describing presence/absence of deterioration. A Poisson fixed effects regression model was applied to an ordered outcome describing the level of deterioration. In both types of models, potential dependency in repeated water measurements was accounted for using household number as the fixed effects variable.

Model results were presented as odds ratios (OR) and risk ratios (RR) for the logistic and Poisson regressions, respectively, along with their 95 percent confidence intervals (CI95%). Overall fit was assessed based on percent of variability explained by the model, calculated from null and residual deviances [VE = (null deviance − residual deviance)/null deviance*100%] and AIC and BIC; and all models were checked to ensure that there were no violations to model assumptions (data not shown).

### 2.3. Association between diarrhea and deterioration

During the study period, weekly diarrhea surveillance was conducted via a household survey (for details see Kattula et al., 2015). Diarrhea outcome was defined as an ordered integer representing the total number of cases per household, and as a binary outcome indicating presence or absence of diarrheal cases in a household during the study period. Since the outcome was aggregated over the study period, for this analysis we removed the temporal component of the exposure and defined tap-to-household deterioration on a household basis. If more than one pair of water samples was available for a household during the study period, we calculated the average of the tap-to-household deterioration for all three spatial approaches. Stratified by presence or absence of diarrhea, we computed descriptive statistics for tap-to-household deterioration, household and tap fecal coliform concentrations, exposure score, crowding, and number of people per age category and compared continuous variables using ANOVA and categorical variables using Pearson's chi-squared test (Supplementary material, Table S2).

We used two types of regression models to investigate the association between diarrheal episodes and tap-to-household deterioration as assessed for each spatial approach (A, B, and C) and accounting for household characteristics. A logistic multivariate regression model was applied to a binary outcome describing presence/absence of diarrheal episodes. A Poisson multivariate regression model was applied to the outcome describing the diarrheal disease counts. Source tap fecal coliform concentration ($\log_{10}$-transformed), crowding and household size were included as additional covariates. All models were checked to ensure that there were no violations to standard model assumptions (data not shown). R statistical software (version 3.1.2) was used for all data processing and statistical analysis.

## 3. Results

### 3.1. Characterization of tap-to-household deterioration

In the study area, 85 out of 160 (53%) study households had water samples collected on the same day from 34 out of 36 (89%) potential source taps (Fig. 2). The number selected as source taps differed with the selected spatial approach: 19 (59%) were assigned as source taps with minimal Euclidean distances (spatial approach A), 20 (63%) with minimal network distances (spatial approach B), and 32 (100%) with inverse network-distance weighted averages (spatial approach C; Table 1).

Fecal coliform concentrations in household water samples ranged from 2 to 6000 CFU/100 ml. The maximum fecal coliform concentration observed in any source tap was 2970 CFU/100 ml. Regardless of spatial approach, selected source taps had on average lower fecal coliform concentrations than households (A: 513±813, B: 484±804, C: 530±779 CFU/100 ml as compared to a household average of 763±1456 CFU/100 ml). Approximately 60% of samples were classified as showing deterioration, meaning the fecal coliform concentrations in household containers were higher than in most of sampled source taps (Table 1; Fig. 2; Supplemental material, Fig. S2). Approximately 14% of samples did not show consistency when being classified as showing signs of deterioration; meaning that those samples were classified as showing deterioration in one approach but not in another. Levels of tap-to-household deterioration ranged from 0 to 3 with an average of 0.71±0.77, 0.73±0.79, and 0.64±0.74 for approaches A, B, and C, respectively. The three spatial approaches produced similar results in terms of frequency and level of tap-to-household deterioration (Table 1).

### 3.2. Association between tap-to-household deterioration and household characteristics

In exploring factors influencing deterioration we found that regardless of spatial approach, the average exposure score was almost two times higher in households with detected deteriorations as compared to households without detected deterioration (∼0.3 vs ∼0.6, $p < 0.02$; Table 2). Spatial approaches that used network distances (B and C), showed that 50% less samples had deterioration when collected during Chill/Cold season (∼20 vs 10, $p < 0.05$, Pearson's chi squared test) than during other seasons. Crowding and size of household did not differ among deterioration categories regardless of the selected spatial approach.

The logistic and Poisson fixed effect models confirmed that deterioration is less likely during the cold season regardless of the selected spatial approach (Table 3). For example, for the weighted network distance approach, water samples collected during Tamil Chill/Cold season were 69% (OR CI95%: 0.11, 0.76; $p = 0.014$) less likely to show tap-to-household deterioration as compared to other seasons. High exposure score could potentially increase the likelihood of deterioration; however the associations were not stable for all spatial approaches (Table 3). For example, only the weighted network distance approach showed that the exposure score was significantly higher in households with detected deterioration: absence of water purification and presence of animals in the home were likely to increase the relative risk of deterioration by 50% ($p < 0.033$).

### 3.3. Association between diarrhea and deterioration

Depending on the spatial approach (A, B, or C), the percentage of households that report diarrhea cases between households that do and do not experience water quality deterioration may differ (see Fig. 2; Supplemental material, Table S2). The weighted network distance approach indicated that households with deterioration are more likely to report cases of diarrhea ($p = 0.039$, Pearson's chi-squared test, Fig. 3). Furthermore, after adjusting for covariates, water quality deterioration based on the weighted network distance approach is likely to triple the chances of reporting diarrhea (OR = 3.08 [1.21, 8.18], $p = 0.02$), Table 4). Although all Poison-based models indicated similar point estimates for water quality deterioration and counts of diarrhea, none of them reached sufficient significance level, yet the number of reported cases of diarrhea increased by approximately 20% with an additional room occupant ($p < 0.01$) consistently across all models.

## 4. Discussion

The present study was conducted in semi-urban slum communities in India with a highly contaminated piped water supply (median FC concentration 80 CFU/100 ml; Kulinkina et al., 2016). A recent review of intermittent water supplies (Kumpel and Nelson, 2016) showed that South Asia has the lowest average number of hours of supply per day (7.2 hours). The frequency of water provision to the study communities is much lower than this average, ranging from once every 2 days to once every 28 days depending on seasonal water availability. In characterizing tap-to-household water quality deterioration under different spatial assumptions, we observed that although source water in the two semi-urban communities is highly contaminated, there is still a high risk of further deterioration during household storage. On average, fecal coliform concentrations were two times higher in household storage containers than in the source taps. This emphasizes the need for household water treatment in addition to improving water quality at the source.

In our study, only the weighted average approach detected a higher risk of water quality deterioration for households that do not purify water and have animals in the home. There are numerous characteristics that might affect water quality in household storage containers including water handling practices, hygiene, and environmental parameters (Brick et al., 2004; Eshcol et al., 2009; Kattula et al., 2015; Levy et al., 2008; Mintz et al., 1995), all of which were not quantified in the present study. The exposure score is not meant to be an exhaustive representation of household characteristics that may affect water deterioration, but rather is an example for illustrating variations in methodology.

Wet season (the Tamil Chill/Cold season) was associated with a lower likelihood of water quality deterioration as compared to other seasons. Although a recent review (Kostyla et al., 2015) showed that levels of microbial contamination in improved drinking water sources are higher during wet conditions, our study showed that the odds of additional contamination from source to household in the study communities are generally lower during the rainy season. This may be caused by higher frequency of water provision and shorter storage times and/or better hygiene associated

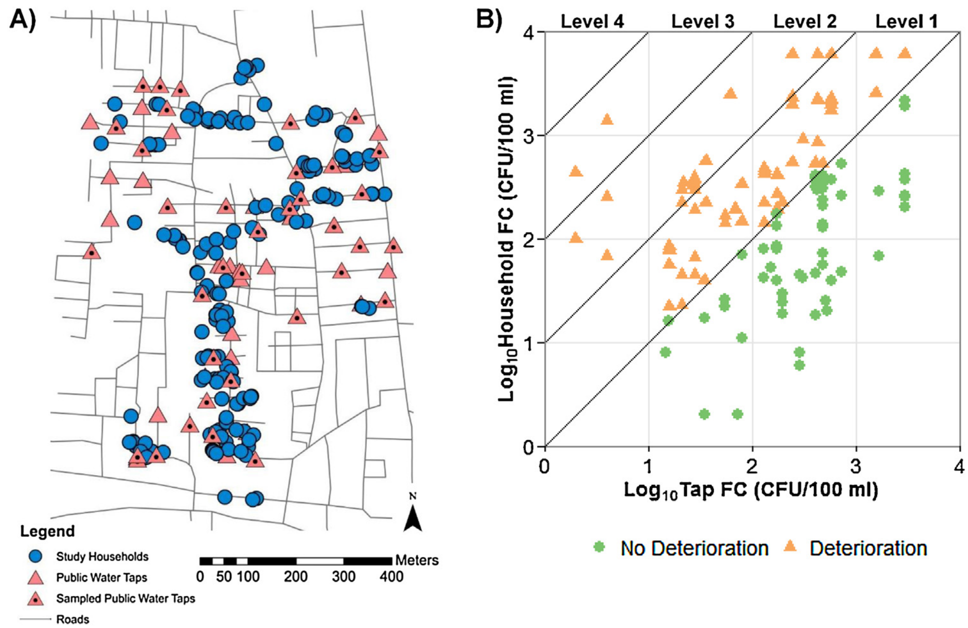


**Fig. 2.** (A) Map of study households and public water taps, Vellore District, India. (B) Tap-to-household water quality deterioration levels (based on fecal coliform [FC] concentrations) for spatial approach A (for B and C see Supplemental Material, Fig. S2).

**Table 1**
Tap-to-household water quality (WQ) deterioration outcome and tap fecal coliform (FC) concentration based on three spatial approaches.

| | Spatial Approach (n = 119)[a] | | | p-value[b] | | |
|---|---|---|---|---|---|---|
| | A: Minimum Euclidean distance | B: Minimum network distance | C: Inverse network distance weighted average | A–B [c] | B–C [d] | A–C [e] |
| **Source Taps** n (%) | 19 (59) | 20 (63) | 32 (100) | 1.000 | **<0.001**[*] | **<0.001**[*] |
| **Tap FC (CFU/100 ml)** Mean (SD) | 513 (813) | 484 (804) | 530 (779) | 0.782 | 0.654 | 0.869 |
| **WQ Deterioration** n (%) | 63 (53) | 64 (54) | 59 (50) | 1.000 | 0.604 | 0.697 |
| **WQ Deterioration Level** Mean (SD) | 0.71 (0.77) | 0.73 (0.79) | 0.64 (0.74) | 0.804 | 0.354 | 0.495 |

*Notes*:
Test between: [c] A and B; [d] B and C; [e] A and C.
[a] 85 out of 160 (53%) study households had water samples collected on the same day as a potential source tap. Household FC = 763 ± 1456 CFU/100 ml.
[b] p-values were calculated using ANOVA for continuous variables and Pearson's chi-squared test for categorical variables.
[*] Statistically significant ($p < 0.05$).

**Table 2**
Detected tap-to-household water quality (WQ) deterioration events based on three spatial approaches.

| | Spatial Approach A | | | Spatial Approach B | | | Spatial Approach C | | |
|---|---|---|---|---|---|---|---|---|---|
| | WQ Deterioration | | | WQ Deterioration | | | WQ Deterioration | | |
| | No (n = 56) | Yes (n = 63) | p-value[a] | No (n = 55) | Yes (n = 64) | p-value[a] | No (n = 60) | Yes (n = 59) | p-value[a] |
| **Exposure Score** Mean (SD) | 0.27 (0.49) | 0.56 (0.64) | **0.007**[*] | 0.27 (0.49) | 0.55 (0.64) | **0.011**[*] | 0.30 (0.50) | 0.54 (0.65) | **0.024**[*] |
| **Tamil Chill/Cold Season** n (%) | 18 (32) | 10 (16) | 0.061 | 18 (33) | 10 (16) | **0.048**[*] | 20 (33) | 8 (14) | **0.020**[*] |
| **Crowding (people/room)** Mean (SD) | 3.56 (1.54) | 4.00 (1.50) | 0.119 | 3.59 (1.51) | 3.97 (1.53) | 0.187 | 3.64 (1.55) | 3.95 (1.49) | 0.273 |
| **Household size (people)** Mean (SD) | 4.98 (1.50) | 5.62 (1.89) | **0.045**[*] | 4.98 (1.52) | 5.61 (1.87) | **0.049**[*] | 5.07 (1.54) | 5.58 (1.90) | 0.110 |
| *Young Children (<5)* | 1.20 (0.44) | 1.25 (0.44) | 0.479 | 1.20 (0.45) | 1.25 (0.44) | 0.539 | 1.20 (0.44) | 1.25 (0.44) | 0.504 |
| *Children (5–15)* | 0.89 (1.06) | 1.06 (1.16) | 0.406 | 0.87 (1.07) | 1.08 (1.15) | 0.317 | 0.85 (1.05) | 1.12 (1.16) | 0.189 |
| *Young Adults (16–40)* | 2.07 (0.76) | 2.33 (0.86) | 0.083 | 2.09 (0.75) | 2.31 (0.87) | 0.143 | 2.17 (0.94) | 2.25 (0.68) | 0.563 |
| *Adults (41–60)* | 0.63 (0.73) | 0.75 (0.67) | 0.347 | 0.64 (0.73) | 0.73 (0.67) | 0.447 | 0.67 (0.71) | 0.71 (0.70) | 0.726 |
| *Elderly (>60)* | 0.20 (0.48) | 0.22 (0.49) | 0.773 | 0.18 (0.47) | 0.23 (0.50) | 0.558 | 0.18 (0.47) | 0.24 (0.50) | 0.546 |

*Notes*:
[a] p-values were calculated using ANOVA for continuous variables and Pearson's chi-squared test for categorical variables.
[*] Statistically significant ($p < 0.05$).

**Table 3**
Relationship between tap-to-household water quality deterioration and exposure score, household size, and Tamil season based on three spatial approaches.

| | Spatial Approach A | | Spatial Approach B | | Spatial Approach C | |
|---|---|---|---|---|---|---|
| | Risk [a] (95% CI) | p-value | Risk [a] (95% CI) | p-value | Risk [a] (95% CI) | p-value |
| *Logistic fixed-effects regression model* | | | | | | |
| **Exposure Score** | 2.28 (1.11, 5.51) | **0.036**[*] | 2.19 (1.05, 5.30) | **0.047**[*] | 1.96 (0.97, 4.24) | 0.064 |
| **Household size** (people) | 1.17 (0.92, 1.58) | 0.226 | 1.18 (0.91, 1.58) | 0.224 | 1.12 (0.88, 1.44) | 0.378 |
| **Tamil Chill/Cold Season** | 0.39 (0.14, 0.95) | **0.048**[*] | 0.37 (0.13, 0.91) | **0.038**[*] | 0.31 (0.11, 0.76) | **0.014**[*] |
| *Poisson fixed-effects regression model* | | | | | | |
| **Exposure Score** | 1.38 (0.96, 1.96) | 0.075 | 1.41 (0.99, 1.99) | 0.056 | 1.50 (1.03, 2.18) | **0.033**[*] |
| **Household size** (people) | 1.08 (0.96, 1.21) | 0.167 | 1.07 (0.96, 1.20) | 0.217 | 1.03 (0.91, 1.17) | 0.601 |
| **Tamil Chill/Cold Season** | 0.51 (0.25, 0.92) | **0.036**[*] | 0.49 (0.24, 0.88) | **0.026**[*] | 0.39 (0.17, 0.77) | **0.012**[*] |

*Notes*:
[a] Odds ratio (OR) for logistic fixed-effects regression model and relative risk (RR) for Poisson fixed-effects regression model.
[*] Statistically significant ($p < 0.05$).

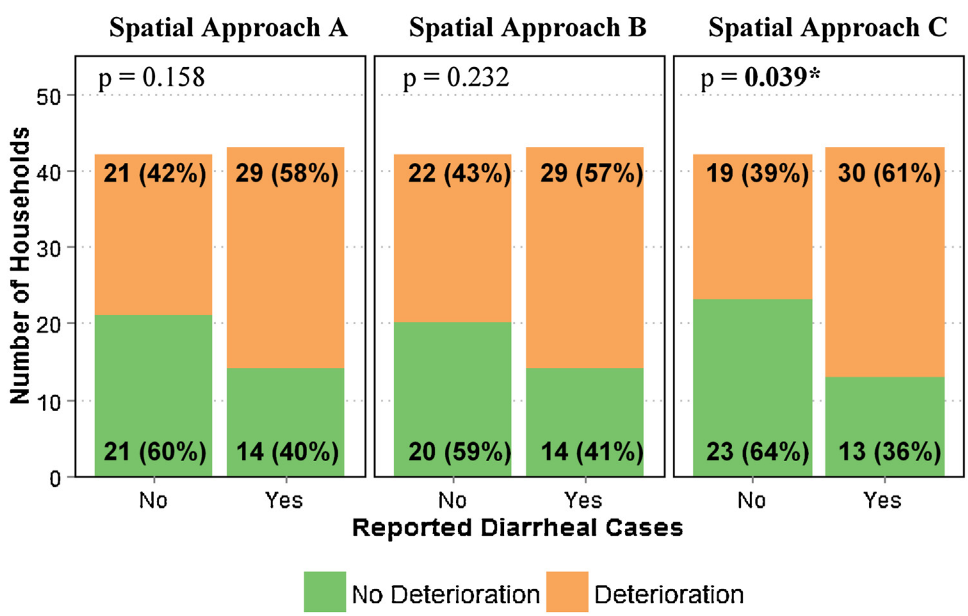


**Fig. 3.** Average household-level reported diarrheal cases and tap-to-household water quality deterioration events based on three spatial approaches. p-values were calculated using Pearson's chi-squared test (* statistically significant; $p < 0.05$).

**Table 4**
Relationship between average household-level reported diarrheal cases and tap-to-household water quality (WQ) deterioration, exposure score, and crowding based on three spatial approaches.

| | Spatial Approach A | | Spatial Approach B | | Spatial Approach C | |
|---|---|---|---|---|---|---|
| | Risk [a] (95% CI) | p-value | Risk [a] (95% CI) | p-value | Risk [a] (95% CI) | p-value |
| *Logistic regression model* | | | | | | |
| **WQ Deterioration** | 2.28 (0.90, 5.96) | 0.085 | 2.06 (0.82, 5.33) | 0.129 | 3.08 (1.21, 8.18) | **0.020**[*] |
| **Exposure Score** | 0.58 (0.25, 1.28) | 0.182 | 0.59 (0.26, 1.31) | 0.205 | 0.55 (0.24, 1.23) | 0.152 |
| **Crowding** (people/room) | 1.22 (0.90, 1.68) | 0.203 | 1.23 (0.91, 1.69) | 0.181 | 1.20 (0.88, 1.65) | 0.255 |
| *Poisson regression model* | | | | | | |
| **WQ Deterioration** | 1.47 (0.95, 2.33) | 0.091 | 1.41 (0.91, 2.23) | 0.133 | 1.44 (0.93, 2.27) | 0.110 |
| **Exposure Score** | 0.78 (0.54, 1.11) | 0.184 | 0.79 (0.54, 1.12) | 0.202 | 0.79 (0.54, 1.12) | 0.204 |
| **Crowding** (people/room) | 1.19 (1.04, 1.35) | **0.011**[*] | 1.19 (1.04, 1.36) | **0.009**[*] | 1.18 (1.04, 1.35) | **0.013**[*] |

*Notes*:
[a] Odds ratio (OR) for logistic regression model and relative risk (RR) for Poisson regression model.
[*] Statistically significant ($p < 0.05$).

with higher water quantity. Our findings thus highlight the need to consider both water quantity and quality when aiming to improve municipal water supplies in urban slums in India.

In our study, we highlight the need to analyze paired water samples in order to truly understand the factors that contribute to water contamination from source to household. However, controlled experiments of this nature such as those conducted by Levy et al. (2008) are rare and may be prohibitively expensive in developing countries, particularly in the context of routine water quality monitoring (Crocker and Bartram, 2014). In a recent meta-analysis, Shields et al. (2015) note that although some studies of changes in water contamination link points between collection and consumption, those links may not represent actual connections for individual samples. The evaluation of various methods of sample pairing we present here is valuable in this context and can inform *post hoc* sample pairing for secondary data analysis. We found that using a distance-weighted average approach, rather than a minimum distance approach, to form links between samples allows for inclusion of a higher number of available samples as well as reduction in the inter-sample variability exhibited by microbial water quality data.

In analyzing the association between diarrheal episodes and tap-to-household water quality deterioration, only the weighted-average method detected a statistically significant positive effect of water quality deterioration on the odds of a household reporting diarrhea cases. Although no association was observed between exposure score and reporting of diarrheal episodes, studies have shown that places with poor environmental sanitation have high burden of diarrhea regardless of drinking water quality (Levy, 2015; Zwane and Kremer, 2007). The location of our study site, Vellore district in Tamil Nadu, India, has been found to have highly contaminated water sources and poor environmental sanitation, including open defecation, presence of garbage in open pits, and high fly density (Collinet-Adler et al., 2015; Sarkar et al., 2013a, 2013b). In such a highly contaminated environment, the risk of water deterioration and diarrhea is likely to continue to be high even if water quality at the source were improved.

The methods we discussed in this study require geocoded water samples which are not always available in developing countries. It should be noted that while we tested multiple spatial assumptions that can be applied to geocoded data, our study still suffers from the limitation of not knowing household water storage times (i.e. temporally linking same day samples from taps and households). Temporal links and changes in associations could be explored using data with higher temporal resolution than what was available for our study.

## 5. Conclusion

Studies of linked water samples and specific attributes of increased contamination between source and household are rare due to methodological challenges and high costs associated with collecting paired samples. Our study demonstrated it is possible to derive useful spatial links between samples *post hoc*; and that the pairing approach affects the conclusions related to associations between enteric infections and water quality deterioration. Our methods may be particularly valuable for developing countries where water quality monitoring is currently limited due to high cost and low human and laboratory capacities (Crocker and Bartram, 2014). Our study encourages spatial analysis on secondary geocoded data to gain deeper knowledge of risk factors of water contamination at the household level and subsequent diarrheal disease risk.

## Conflict of interest

The authors declare they have no actual or potential competing financial interests.

## Acknowledgments

This research was supported by the Indian Council of Medical Research and Centers for Disease Control and Prevention through the Indo US CDC/ICMR grant (IRIS ID 2008-07980), and Tufts Institute for Innovation. T.M.A. was supported by a scholarship from Tufts University School of Engineering.

CMC: Mr. Senthil, Mr. Arul Raj, Mrs. Bhuvaneshwari, Ms. Suganthi, Mr. Srinivasan and Mr. Karuna for data collection and entry; Wellcome Research Laboratory for conducting laboratory testing; study families for their participation.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijheh.2016.09.019.

## References

Abdellah, A.M., Abdel-Magid, H.M., Yahia, N.A., 2012. Assessment of drinking water microbial contamination in Al-Butana region of Sudan. J. Appl. Sci. 12, 856–862, http://dx.doi.org/10.3923/jas.2012.856.862.

Aldana, J.M., 2010. Rapid Assessment of Drinking-Water Quality in the Republic of Nicaragua: Country Report of the Pilot Project Implementation in 2004–2005. World Health Organization/United Nations Children's Fund, Geneva.

Aliev, S., Shodmonov, P., Babakhanova, N., Schmoll, O., 2010. Rapid Assessment of Drinking-Water Quality in the Republic of Tajikistan: Country Report of the Pilot Project Implementation in 2004–2005. World Health Organization/United Nations Children's Fund, Geneva.

Bain, R., Cronk, R., Hossain, R., Bonjour, S., Onda, K., Wright, J., et al., 2014a. Global assessment of exposure to faecal contamination through drinking water based on a systematic review. Trop. Med. Int. Health 19 (8), 917–927, http://dx.doi.org/10.1111/tmi.12334.

Bain, R., Cronk, R., Wright, J., Yang, H., Slaymaker, T., Bartram, J., 2014b. Fecal contamination of drinking-water in low- and middle-income countries: a systematic review and meta-analysis. PLoS Med. 11 (5), e1001644, http://dx.doi.org/10.1371/journal.pmed.1001644.

Bhunia, R., Ramakrishnan, R., Hutin, Y., Gupte, M.D., 2009. Cholera outbreak secondary to contaminated pipe water in an urban area, West Bengal, India, 2006. Indian J. Gastroenterol. 28 (2), 62–64, http://dx.doi.org/10.1007/s12664-009-0020-5.

Biswas, A.K., 2007. New Insights on Water Security in Asia. In: Asian Water Development Outlook 2007: Achieving Water Security for Asia. Asian Development Bank, pp. 1–40.

Brick, T., Primrose, B., Chandrasekhar, R., Roya, S., Muliyilb, J., Kanga, G., 2004. Water contamination in urban south India: household storage practices and their implications for water safety and enteric infections. Int. J. Hyg. Environ. Health 207 (5), 473–480, http://dx.doi.org/10.1078/1438-4639-00318.

Collinet-Adler, S., Babji, S., Francis, M., Kattula, D., Premkumar, P.S., Sarkar, R., et al., 2015. Environmental factors associated with high fly densities and diarrhea in Vellore, India. Appl. Environ. Microbiol. 81 (17), 6053–6058, http://dx.doi.org/10.1128/AEM.01236-15.

Crocker, J., Bartram, J., 2014. Comparison and cost analysis of drinking water quality monitoring requirements versus practice in seven developing countries. Int. J. Environ. Res. Public Health 11 (7), 7333–7346, http://dx.doi.org/10.3390/ijerph110707333.

Eshcol, J., Mahapatra, P., Keshapagu, S., 2009. Is fecal contamination of drinking water after collection associated with household water handling and hygiene practices? A study of urban slum households in Hyderabad, India. J. Water Health 7, 145–154.

Firth, J., Balraj, V., Muliyil, J., Roy, S., Rani, L.M., Chandresekhar, R., et al., 2010. Point-of-use interventions to decrease contamination of drinking water: a randomized, controlled pilot study on efficacy, effectiveness, and acceptability of closed containers, Moringa oleifera, and in-home chlorination in rural South India. Am. J. Trop. Med. Hyg. 82 (5), 759–765, http://dx.doi.org/10.4269/ajtmh.2010.09-0206.

Kattula, D., Francis, M.R., Kulinkina, A., Sarkar, R., Mohan, V.R., Babji, D., et al., 2015. Environmental predictors of diarrhoeal infection for rural and urban communities in south India in children and adults. Epidemiol. Infect. 143 (14), 3036–3047, http://dx.doi.org/10.1017/S0950268814003562.

Kostyla, C., Bain, R., Cronk, R., Bartram, J., 2015. Seasonal variation of fecal contamination in drinking water sources in developing countries: a systematic review. Sci. Tot. Environ. 514, 333–343;, http://dx.doi.org/10.1016/j.scitotenv.2015.01.018.

Kulinkina, A.V., Mohan, V.R., Francis, M.R., Kattula, D., Sarkar, R., Plummer, J.D., et al., 2016. Seasonality of water quality and diarrheal disease counts in urban and rural settings in south India. Sci. Rep. 6, 20521, http://dx.doi.org/10.1038/srep20521.

Kumpel, E., Nelson, K.L., 2016. Intermittent water supply: prevalence, practice, and microbial water quality. Environ. Sci. Technol. 50 (2), 542–553, http://dx.doi.org/10.1021/acs.est.5b03973.

Levy, K., 2015. Does poor water quality cause diarrheal disease? Am. J. Trop. Med. Hyg. 93 (5), 899–900, http://dx.doi.org/10.4269/ajtmh.15-0689.

Levy, K., Nelson, K.L., Hubbard, A., 2008. Following the water: a controlled study of drinking water storage in northern coastal Ecuador. Environ. Health Perspect. 116 (11), 1533–1540, http://dx.doi.org/10.1289/ehp.11296.

Mintz, E.D., Reiff, F.M., Tauxe, R.V., 1995. Safe water treatment and storage in the home: a practical new strategy to prevent waterborne disease. JAMA 273, 948–953.

Prüss-Ustün, A., Bartram, J., Clasen, T., et al., 2014. Burden of disease from inadequate water, sanitation and hygiene in low- and middleincome settings:

a retrospective analysis of data from 145 countries. Trop. Med. Int. Health 19, 894–905.

Sarkar, R., Prabhakar, A.T., Manickam, S., Selvapandian, D., Raghava, N.V., Kang, G., et al., 2007. Epidemiological investigation of an outbreak of acute diarrhoeal disease using geographic information systems. Trans. R. Soc. Trop. Med. Hyg. 101 (6), 587–593, http://dx.doi.org/10.1016/j.trstmh.2006.11.005.

Sarkar, R., Ajjampur, S.S., Prabakaran, A.D., Geetha, J.C., Sowmyanarayanan, T.V., Kane, A., et al., 2013a. Cryptosporidiosis among children in an endemic semiurban community in southern India: does a protected drinking water source decrease infection? Clin. Infect. Dis. 57 (3), 398–406, http://dx.doi.org/10.1093/cid/cit288.

Sarkar, R., Sivarathinaswamy, P., Thangaraj, B., Sindhu, K.N., Ajjampur, S.S., Muliyil, J., et al., 2013b. Burden of childhood diseases and malnutrition in a semi-urban slum in southern India. BMC Public Health 13 (87), http://dx.doi.org/10.1186/1471-2458-13-87.

Shields, K.F., Bain, R.E., Cronk, R., Wright, J.A., Bartram, J., 2015. Association of supply type with fecal contamination of source water and household stored drinking water in developing countries: a bivariate meta-analysis. Environ. Health Perspect. 123, 1222–1231, http://dx.doi.org/10.1289/ehp.1409002.

Trevett, A.F., Carter, R.C., Tyrrel, S.F., 2005. The importance of domestic water quality management in the context of faecal-oral disease transmission. J. Water Health 3 (3), 259–270, http://dx.doi.org/10.2166/wh.2005.037.

UNICEF/FAO/SaciWATERs, 2013. Water in India: situation and prospects. UNICEF, FAO, SaciWATERs, New Delhi.

WHO, 2011. Technical Guidance on Water-related Disease Surveillance. WHO, Geneva, file/0009/149184/e95620.pdf (accessed 30.10.14.).

WHO/UNICEF, 2014. Progress on Drinking-Water and Sanitation: 2014 Update. WHO/UNICEF, Geneva.

Wright, J., Gundry, S., Conroy, R., 2004. Household drinking water in developing countries: a systematic review of microbiological contamination between source and point-of-use. Trop. Med. Int. Health 9, 106–117.

Zwane AP, and Kremer M., What works in fighting diarrheal diseases in developing countries? A critical review. CID Working Paper No. 140. March 2007.