Refined Assessment of Associations between Drinking Water Residence Time and Emergency Department Visits for Gastrointestinal Illness in Metro Atlanta, Georgia

Karen Levy¹,²,*, Mitchell Klein¹, Stefanie Ebel Sarnat¹, Samina Panwahar²,³,**, Alexandra Huttinger²,³, Paige Tolbert¹, and Christine Moe²,³

¹Department of Environmental Health, Rollins School of Public Health, Emory University, 1518 Clifton Rd NE, Atlanta, GA, 30322
²Center for Global Safe Water, Sanitation and Hygiene at Emory University, 1518 Clifton Rd NE, Atlanta, GA, 30322
³Hubert Department of Global Health, Rollins School of Public Health, Emory University, 1518 Clifton Rd NE, Atlanta, GA, 30322

Abstract

Recent outbreak investigations suggest that a substantial proportion of waterborne disease outbreaks are attributable to water distribution system issues. In this analysis, we examine the relationship between modeled water residence time (WRT), a proxy for probability of microorganism intrusion into the distribution system, and emergency department visits for gastrointestinal (GI) illness for two water utilities in Metro Atlanta, USA during 1993–2004. We also examine the association between proximity to the nearest distribution system node, based on patients’ residential address, and GI illness using logistic regression models. Comparing long (≥90th percentile) to intermediate WRTs (11th to 89th percentile), we observed a modestly increased risk for GI illness for Utility 1 (OR=1.07, 95% CI: 1.02–1.13), which had substantially higher average WRT than Utility 2, for which we found no increased risk (OR=0.98, 95% CI: 0.94–1.02). Examining finer, 12-hour increments of WRT, we found that exposures >48 hrs were associated with increased risk of GI illness, and exposures of >96 hrs had the strongest associations, although none of these associations were statistically significant. Our results suggest that utilities might consider reducing WRTs to <2–3 days or add booster disinfection in areas with longer WRT, to minimize risk of GI illness from water consumption.

Keywords

Water distribution system; water quality; emergency department; gastrointestinal illness; diarrhea

*Corresponding author. Phone: 404-727-4502; Fax: 727-8744; karen.levy@emory.edu.
**Current Address: Oregon Health Authority, 500 Summer St. NE, E52, Salem, OR 97301
Introduction

The aging water delivery infrastructure is becoming a growing issue in the United States, with deferred maintenance creating a system that is “deeply stressed,” “over-worked and under-budgeted”, according to the U.S. Environmental Protection Agency (USEPA 2011a). There are an estimated 240,000 water main breaks per year in the U.S., and the number of breaks increases substantially near the end of the system's service life (USEPA 2011b). Pipe breaks and/or low and negative pressure events in the drinking water distribution system can result in intrusion of pathogenic microorganisms if an external source of contamination is present (Besner et al. 2011).

Several outbreaks of waterborne disease have been attributed to public drinking water systems, including the largest reported waterborne disease outbreak ever documented, in Milwaukee, Wisconsin in 1993, with over 400,000 people affected (MacKenzie et al. 1995). Between 1971 and 2006, 282 outbreaks of acute gastrointestinal (GI) illness in community water systems were reported to the U.S. Waterborne Disease and Outbreak Surveillance System (WBDSS). Contamination within the distribution system and in- premise plumbing accounted for 79 (9.9%) and 65 (8.1%) respectively, of the non-legionellosis waterborne disease outbreaks where a deficiency in the water system could be identified. While there has been a decrease in the number of reported outbreaks over time, in this period there was no change in the annual proportion of outbreaks associated with distribution system deficiencies (Craun et al. 2010). Between 2009–2010, distribution system deficiencies accounted for 15.2% (5/33) of identified outbreak deficiencies in drinking water-associated disease outbreaks (CDC 2013).

However, outbreaks of waterborne disease represent only the disease events that impact a large enough number of people within a relatively short period of time to be recognized by public health agencies. Linking illness to drinking water is inherently difficult through outbreak investigation methods because most persons have daily exposure to tap water (Tostmann et al. 2012; CDC 2013). Epidemiological studies can be used to evaluate the endemic burden of waterborne disease in the population, and to examine risks associated with distribution system deficiencies and intermittent contamination. A number of observational studies have found associations between GI illness and drinking tap water compromised by low pressure events (Hunter et al. 2005; UK), main breaks or maintenance work (Nygard et al. 2007; Norway), declines of residual chlorine concentrations (Egorov et al. 2002; Russia), increased water age (Tinker et al. 2009; USA), increased water turbidity (Egorov et al. 2003; Russia; Tinker et al. 2010; USA), water system outages (CDC 2011; USA), and virus detection in non-disinfected groundwater (Borchardt et al. 2012; USA). On the other hand, Malm et al. (2013; Sweden) found no evidence of increased complaints of GI illness associated with disturbances at the water works or in the distribution network compared to control periods without disturbances in the same geographical area, in a study using Swedish national Health Call Center data. A recent review of the impact of distribution system deficiencies on endemic gastrointestinal illness found that tap water consumption in malfunctioning distribution networks and system deficiencies, such as water outages, were both associated with gastrointestinal disease (Ercumen et al. 2014).
assessment models also suggest that distributions system can be a source of GI illness (Teunis et al. 2010; USA; Lambertini et al. 2012; USA).

Intervention studies comparing homes with water treatment devices versus those without have had mixed results, with some reporting increased risk of GI illness attributable to drinking tap water (Payment et al. 1991, Payment et al. 1997; Canada), some reporting increased risk only in sensitive sub-populations (Colford et al. 2009; USA), and others reporting no difference (Hellard et al. 2001; Australia; Colford et al. 2005; USA). In their meta-analysis, Ercumen et al. (2014) found elevated risk of gastrointestinal illness for consumers of tap water versus point-of-use treated water in unblinded studies, but no association for studies that blinded participants to their point-of-use water treatment status.

Epidemiological studies of distribution system contributions to GI illness are challenging to design and implement because of the difficulty in estimating exposure to waterborne pathogens, detecting the relevant health outcomes, and controlling for the effect of confounding factors. Our group has taken advantage of an extensive dataset of emergency department (ED) visits and detailed information on water distribution systems in Metro Atlanta, GA to address this challenge. We previously conducted an analysis of the associations between ED visits for acute GI illness and water residence time (WRT) in the distribution system, as estimated using hydraulic models developed by two water utilities. We consider WRT as a proxy for microorganism intrusion into the distribution system, because when it takes longer for water to reach the consumer there is a higher probability of an intrusion event occurring. As hypothesized, we observed a modest increased risk for GI illness, approximately five to seven percent, among people living in ZIP codes served by water with a long average WRT (top 10 percent) compared to intermediate residence times (11th to 89th percentile), after controlling for potential confounding factors such as patient age and markers of socioeconomic status (Tinker et al. 2009).

In this previous analysis, we based the WRT exposure assignment on a ZIP code average level. However, important variation in WRT likely exists within ZIP codes as the lengths of water pipes between the fixed-location water treatment plants and homes vary, depending on size of ZIP code and location of homes. Because we are ultimately interested in the water that reaches the end user, we hypothesized that a more spatially refined characterization of WRT reduces exposure misclassification and thus may reduce any bias to the null of our previous epidemiologic findings. Therefore, in the analysis presented here, we refined our exposure metric by estimating patient-level WRT based on patients’ residential address and proximity to the nearest distribution system node in a calibrated hydraulic model of the distribution system.

**Methods**

**Study Site**

The greater Atlanta metropolitan area is served by six major water utilities. The water treatment technology, age of the infrastructure, size of the service area, demographics of the population served, and various other factors differ amongst these utilities. In this way, Atlanta provides a good example of the range of conditions that exist in drinking water
distribution systems in major U.S. cities. This study focuses on two utilities in the greater Atlanta area. Utility 1 and Utility 2 each operate two treatment plants that supply water to their respective distribution pipeline networks. Utility 1 serves 680,000 customers over a 650 square mile area, and Utility 2 serves 1.2 million customers over a 348 square mile area. The demographics of the populations served differ substantially: Utility 1 serves a 65% Caucasian population with an average Census 2000 block-level median household income of $60,569 (s.d. $20,017), whereas Utility 2 serves an 85% African American population with an average Census 2000 block-level median household income of $28,894 (s.d. $16,674).

**Emergency Department Data**

Our ED database was comprised of over 10 million ED visit records from hospitals in the 20-county Atlanta metropolitan area during 1993–2004, in which 41 of 42 acute care hospitals provided data for all or part of the study period. Relevant data elements for the current analysis included: patient medical record number, unique visit number, date of admission, primary and secondary International Classification of Diseases, Ninth Revision (ICD-9) diagnostic codes, patient age, date of birth, gender, race, ZIP code of residence, residential street address, and method of payment for the visit (e.g., Medicaid).

The current analysis was based on a subset of our overall database, and a subset of those records analyzed by Tinker et al (2009). In this analysis, records were selected for inclusion for patients who visited any of the participating hospital EDs during the study period, had full residential address data available, and resided in the service area of the selected water companies at the time of the ED visit. Our initial selection resulted in 1,772,787 records from 15 hospitals for which full residential address information was available and which were made by patients with residential ZIP codes in either the Utility 1 (884,643 visits) or Utility 2 (888,144 visits) service areas, as defined by Tinker et al. (2009). This dataset differs from that used in the Tinker et al. (2009) analysis, which included 2,092,735 records from 27 hospitals, because that analysis was not restricted by a requirement for full residential address information. A flow chart of the data processing steps we followed is shown in Figure 1, and a comparison of the data used in the present analysis to the Tinker et al. (2009) analysis is presented in Table 1.

Our GI illness outcome encompassed ED visits for which the primary, or any available secondary ICD-9, code had one of the following diagnoses: infectious GI illness (001–004, 005.0, 005.4, 005.89, 005.89, 005.9, 006–007, 008.0, 008.42–008.44, 008.47, 008.49, 008.5, 008.6, 008.8, 009), non-infectious GI illness (558.9), and nausea and vomiting plausibly related to GI illness (787.01–787.03, 787.91). Non-infectious GI illness was included in the case definition because previous research has shown that many infectious cases of GI illness are misclassified into this diagnostic category (Gangarosa et al., 1992; Lew et al., 1991; Schwartz, Levin and Hodge, 1997). The control group included non-injury ED visits without GI illness. The spatial distribution of ED usage for injuries might be different than the spatial distribution of the source population, so we excluded ED visits for injuries in the comparison group in an effort to better track the spatial distribution of the source population (i.e., those that would go to the ED for a GI illness if they had a GI illness). Repeat visits by a patient within a single day were counted as a single visit.
Ethics Approval

This study protocol was approved by the Social, Humanist, and Behavioral Committee for the Protection of Human Subjects, by the Institutional Review Board at Emory University, Atlanta, GA, USA (Protocol #IRB00045761).

Water Residence Time Data

WRT was estimated using extended period hydraulic models provided by the two water utilities. These models estimate the water travel times to all pipe intersections ('nodes'), which represents a flow-weighted average time for all travel paths between the treatment plant and the node within the distribution system pipe network. Further details of water age calculations are provided in Tinker et al. (2009). Tinker et al. (2009) utilized a usage-weighted average of the node WRT by ZIP code. Here we used individual node-level WRT estimates. WRT was estimated for Utility 1 from 1999–2003 and Utility 2 from 1993–2004.

Spatial Data Processing

All ED visit records were geocoded using patient residential address information to enable assignment of WRT at the closest node in the water distribution system. Geocoding was conducted using a standard geocoding algorithm in Streetmap for ArcGIS version 9.2 (ESRI, Redlands, CA). Records with invalid address information and observations with a geocoding confidence score of less than 50% were removed (Figure 1).

We joined node-level WRT data, geocoded ED visit data, and census block-level median household income and percent minority data from the 2000 US Census (United States Census Bureau 2000) in ArcMap version 9.3 (ESRI, Redlands, CA), separately for each of the two service areas. The joining function assigned the nearest water distribution system node to each ED patient residence and calculated the Euclidean distance for each pairing. At this step, we excluded ED visits with residences that were located outside of the service area, as defined by geographic boundary data obtained from each utility (Figure 1). This addressed the issue that in addition to serving individual household customers directly, both utilities sell their water to several wholesale city customers that are located within the service areas, but whose water nodes are not included in the hydraulic models.

Statistical Analysis

We used a logistic regression model to estimate the odds of ED patients having GI illness as a function of WRT at the node nearest their residence, comparing ED patients with GI illness (GI=1) to non-injury ED patients without GI illness (GI=0). The basic logistic model had the following form:

$$\text{Logit } P(\text{GI}=1) = \alpha + \sum \beta_i (\text{WRT}_i) + \text{covariates}$$

where WRT$_i$ represents the $i^{th}$ category of WRT.

Separate models were run for Utility 1 and Utility 2. First, WRT was modeled in three categories, with $a priori$ cut points: ≤10th percentile, 10th - 90th percentile, ≥90th percentile.
The middle category (10th - 90th percentile) was the reference category, following Tinker et al. (2009). We also modeled WRT for Utility 2 using the cut points for Utility 1 as a comparison, in order to compare the same absolute levels for WRT decile categorization. Second, for ease of interpretation for water systems engineers, WRT was modeled using absolute number of hours, in nine categories of 12-hour increments, with <12 hrs as the reference category.

Control variables (covariates) included: Medicaid status (patient level indicator–yes, no, or missing), median household income (census block level–continuous), percent minority (census block level–continuous), age (patient level indicator with four levels: 0–5, 6–18, 19–64, ≥65), Euclidean distance between hospital where ED visit occurred and geocoded patient residence (patient level indicator with three levels: <20th, 20–80th, >80th percentile), hospital indicator variables denoting the hospital in which the ED visit occurred, year indicator variables, day-of-week indicator variables, season indicator variable, and product terms between patient age and hospital, patient age and distance to hospital, and patient age and Medicaid status. Control variables were chosen on the basis of being potential risk factors for GI illness that also vary spatially, and thus could also be associated with WRT.

For both utilities, we also used a logistic regression model to compare the odds of ED patients having GI illness as a function of straight-line Euclidian distance between x-y coordinates from households to nodes, controlling for the same set of variables as described above. We compared distances of 100–500 ft and >500 ft to a reference category of <100 ft.

Results

Summary Statistics

Frequencies of visits for gastrointestinal illness among non-injury ED visits during the study periods are shown in Table 1. The geocoding procedure resulted in 698,705 visits with valid geocodes for Utility 1 and 708,330 visits with valid geocodes for Utility 2. This represents an overall success rate of 80.6%. Total numbers of visits and nodes included in the analysis are presented in Table 1. Geocoded data were available for only approximately half of the hospitals and total visits included in the analysis by Tinker et al. (2009). Utility 1 had a significantly higher average WRT than Utility 2 (p<0.0001), with a mean of 55.9 hrs (range: 0.51–336 hrs) versus 16.1 hrs (range: 0.24–336 hrs), as expected given the difference in the size of their service areas.

Epidemiologic Analysis

Figure 2 shows the results of the logistic regression models, comparing the exposure assessment approach used here to the results we previously reported in Tinker et al. (2009). Using geocoded addresses, we found similar results for Utility 1 to those found by Tinker et al. (2009), with a slightly more elevated association between long WRT (≥03.6 hrs) and GI illness (OR=1.07, 95% CI: 1.02–1.13), and no protective effect of short WRT (≤3.8 hrs) and GI illness (OR=1.00, 95% CI: 0.95–1.06). For Utility 2, we did not observe an increased risk of longer WRT (≥29.0 hrs) and GI illness in this analysis (OR=0.98, 95% CI: 0.94–
1.02), nor a protective effect of short WRT (5.8 hrs) and GI illness (OR=1.00, 95% CI: 0.96–1.04).

Because the cutpoint for long WRT was substantially greater for Utility 1 than for Utility 2, and because we observed stronger associations with GI ED visits for Utility 1, we also considered the decile cutpoints for Utility 1 in the analysis for Utility 2. This allowed for a direct comparison of the two utilities relative to the same WRT categorization. This analysis showed elevated risk for longer WRT, but with a very large confidence interval for the long WRT category risk ratio.

Figure 3 shows the results of the logistic regression models evaluating finer increments of WRT. While confidence intervals overlap among estimated effects, the strongest effects were observed at the longest WRT category (>96 hrs) for both utilities, and with a suggestion of an upward trend with step-wise increases in WRT for Utility 2.

We also evaluated the odds of increased risk of GI illness ED visits with Euclidian distance between geocoded patient residences and location of nearest distribution system node (Table 2). For Utility 1, we observe an increased odds of GI illness for patients living >500 ft from the nearest node, compared to a reference category of <100 ft (OR=1.09, 95% CI: 1.03–1.15).

Discussion

Ultimately, exposure to microbial contamination within the distribution system is dependent upon the occurrence of low-pressure conditions, presence of a source of contamination, and existence of a pathway for entry of external contaminants, which together can create the necessary conditions for intrusion events. The concentrations of microbes and duration of intrusion events are key factors influencing exposure (Besner et al. 2011). We use WRT here as a proxy for a large number of processes within the drinking water distribution system that determine whether a person opens their tap at the time when a contaminant is passing through and drinks that water. Our assumption is that longer WRT increases the probability of a number of intrusion events occurring.

We conducted this analysis to explore how a more geographically-resolved exposure assignment might affect the associations between WRT and ED visits for GI illness observed by Tinker et al. (2009), who assigned aggregated WRT by ZIP code. We performed this analysis for two drinking water utilities in the greater Atlanta metropolitan area, with different demographic and distribution system characteristics. In particular, Utility 1 had a substantially higher average WRT as compared to Utility 2.

Comparing long WRT (90th percentile) to intermediate residence times (11th to 89th percentile), we observed a modestly increased risk for GI illness for Utility 1 (OR=1.07, 95% CI: 1.02–1.13) but not for Utility 2 (OR=0.98, 95% CI: 0.94–1.02), whereas Tinker et al. (2009) observed a modestly increased risk for GI illness for both utilities (Utility 1: OR=1.07, 95% CI: 1.03–1.10; Utility 2: OR=1.05, 95% CI: 1.02–1.08). When we compared the same categories of WRT for both utilities (i.e., based on quintiles of WRT at Utility 1), we observed an increased odds of illness but with a large amount of uncertainty in the
estimates (OR=1.25, 95% CI: 0.96–1.64), because of the few WRT observations that occurred in the long WRT category at Utility 2 in this analysis. Fewer than 1% of the Utility 2 observations occurred in the highest 20th percentile of Utility 1 WRTs.

It is important to note that Tinker et al. (2009) made an a priori decision to use the intermediate WRT category as the reference group and to explore the possibility of increased risk related to short WRTs due to reduced contact time with disinfectant in the distribution system, in addition to exploring the possibility of increased risk related to longer WRTs due to increased probability of contamination within the distribution system. In order to maintain consistency between the analyses, we also used the intermediate WRT category as the reference group.

The analysis of 12-hour increments of WRT in relation to ED visits for GI illness provided additional insights that are more interpretable to water treatment plant operators. Examining these finer increments of WRT, we found that only exposures >48 hrs were associated with increased risk of GI illness, and exposures of >96 hrs had the strongest associations (Figure 3), although none of these relationships were statistically significant. While there was very little data for Utility 2 in the longest WRT categories, in both utilities, we observed an upward trend in risk at the longest absolute WRTs. Again, these relationships were not statistically significant.

By comparison, Payment et al. (1991, 1997) reported WRT ranging from 0.3–34 hrs in Montreal, in studies that found no correlation between WRT and highly credible gastrointestinal illness. Hellard et al. (2001) reported average WRTs of 24–36 hrs in Melbourne, where no difference was found between groups with functional versus sham household water treatment devices. Thus, the distribution systems observed in these studies may not have reached values of WRT to cause elevated risk of GI illness. However, because WRT is a proxy for likelihood of pathogen intrusion, and each distribution system is unique, we would not necessarily expect the same absolute values across systems (e.g., an older system may have more intrusions within the same WRT relative to a newer system).

**Exposure Assessment**

This analysis provided the opportunity to compare different approaches to WRT exposure assignment. The node-level WRT used here provided a more refined exposure assignment than WRT aggregated by ZIP code, and provided the opportunity for an individual-level, rather than ecological analysis. Aggregation of WRT estimates across nodes within a ZIP code smooths out heterogenous exposures across an area. Thus, a small number of ZIP codes can have a large influence on the results, and might misrepresent the overall exposure for a particular ZIP code.

However, it is important to note that there were differences in the two datasets analyzed (Table 1), and our refined exposure assignment approach presented several limitations. Assigning node-level WRT to ED patients reduced the sample size available for the analysis as it required limiting the dataset to ED visits with patient residential address information; this reduced the number of hospitals and our total sample size by about half compared to the Tinker et al. (2009) analysis. More specific exposure assignment also opens up the
possibility of errors related to address and node assignment. Assigning WRT to a specific node in the distribution system might assume too much specificity, i.e., the nearest node by Euclidian distance might not in fact be the node from which the household receives its water, due to complexities in the distribution system. In addition, individuals may consume water outside of the home, but still in the same general vicinity, for example at a local school or childcare center. More aggregated exposure assessment averages out these types of errors.

Our exposure assessment was also limited to describing WRT at distribution system nodes, rather than to the tap, at the point of use. The ideal measure of WRT would describe time for water to reach the user at their residence. The availability of geocoded data for visitors to the ED, as well as the locations of nodes, allowed us to examine Euclidian distance between residence and node in the distribution system. We found that for Utility 1, ED visitors living more than 500 ft from the nearest node in the distribution system had 9% increased odds (95% CI: 3–15%) of having GI illness compared to those living less than 100 ft from the nearest node (Table 2). This analysis must be interpreted with caution because Euclidian distance to a distribution system node does not necessarily represent whether a household receives water from that distribution system node. However, distance between households and drinking water delivery nodes might be an area to explore in future investigations.

It is also important to note that we had no data on water consumption patterns, so our analyses relied on the assumption that patients were consuming water directly from the distribution system, and we also had no information on quality of premise plumbing. Consumption of purchased water or point-of-use treatment of distribution system water would therefore affect the results we observed. Additionally, the assignment of WRT at each node is subject to error due to factors such as misspecifications of the hydraulic model, mixing of water of different ages, and assignment of only one WRT measurement per node per year.

Conclusions

Our results suggest that drinking water utilities might consider reducing absolute WRTs to less than 2–3 days or consider adding booster disinfection in areas of the distribution system with longer WRT in order to minimize risk of GI illness from consumption of municipal water. Continued investment is needed to maintain the current quality of drinking water infrastructure.

Acknowledgments

Financial support was provided by United States Environmental Protection Agency (USEPA) Grants #R834250, #R831629, #R82921301, by National Institute of Environmental Health Sciences Grant #RO1ES011294; by National Institute for Allergy and Infectious Diseases Grant #1K01AI103544; and by Electric Power Research Institute Grants #W03253-07, EP-P27723/C13172, EP-P4353/C2124. We would also like to thank the water utility collaborators who provided hydraulic models for the analysis. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication.
References


Figure 1.
Data sources and data processing steps involved in the analysis of the association between emergency department (ED) visits and water residence time for two water utilities in Metro Atlanta, GA, USA.
Figure 2. Comparison of Results of Logistic Regression Models Using Different Approaches to Exposure Assessment

Odds Ratio (OR) estimates and 95% confidence intervals are shown for Utility 1 and Utility 2 for the association between water residence time (WRT) and visits to the emergency department for gastrointestinal illness. WRT was classified as "Short" (≤10th percentile), "Intermediate" (11th–89th percentile), or "Long" (≥90th percentile). Squares indicate OR using WRT aggregated by ZIP code, as reported by Tinker et al. (2009). Circles indicate OR using geocoded WRT, with address of patient matched to WRT of nearest node. Closed circles represent OR using the intermediate WRT category of each utility as its own reference. Open circles represent OR for Utility 2 relative to the WRT category percentile cut-off values of Utility 1. Models controlled for Medicaid status, median household income at the census block level, percent minority at the census block level, patient age, Euclidean distance between hospital where ED visit occurred and geocoded patient residence, indicator variables for hospital, year, day-of-week, season, and product terms for age*hospital, age*distance to hospital, age*Medicaid status.
Figure 3. Results of Logistic Regression Models Examining the Relationship between Water Residence Time (12-hr increments) vs. Emergency Department visits for Gastrointestinal Illness

Odds Ratio (OR) estimates and 95% confidence intervals are shown for a) Utility 1 and b) Utility 2 for the association between water residence time (WRT) in 12-hour increments and visits to the emergency department for gastrointestinal illness, using geocoded WRT estimates, with address of patient matched to WRT of nearest node. Reference condition for each estimate is <12 hrs. Mean WRT for each utility shown with a dotted line. Models controlled for Medicaid status, median household income at the census block level, percent
minority at the census block level, patient age, Euclidean distance between hospital where ED visit occurred and geocoded patient residence, indicator variables for hospital, year, day-of-week, and product terms for age*hospital, age*distance to hospital, age*Medicaid status.
Table 1

Comparison of approaches and data used in the present analysis versus that used by Tinker et al. (2009).
WRT=Water residence time. ED=Emergency department. GI=Gastrointestinal.

<table>
<thead>
<tr>
<th>Exposure Assessment Approach</th>
<th>Tinker et al. (2009)</th>
<th>Present Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WRT aggregated by ZIP code</td>
<td>WRT assigned by matching geocoded address to nearest node in the distribution system</td>
</tr>
<tr>
<td>Differences in Calculation of Control Variables</td>
<td>• Median household income and percent minority aggregated by ZIP code</td>
<td>• Median household income and percent minority by census block</td>
</tr>
<tr>
<td></td>
<td>• Distance from hospital calculated to zip code centroid</td>
<td>• Distance from hospital calculated to geocoded residence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># of Hospitals Included</th>
<th>28</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility 1</td>
<td>Utility 2</td>
<td>Utility 1</td>
</tr>
<tr>
<td># of ED visits analyzed **</td>
<td>721,982</td>
<td>1,205,816</td>
</tr>
<tr>
<td># of GI illness ED visits</td>
<td>63,652</td>
<td>101,285</td>
</tr>
<tr>
<td>% of ED visits for GI illness</td>
<td>8.8%</td>
<td>8.4%</td>
</tr>
<tr>
<td>Mean WRT (Range)</td>
<td>32.8 hrs (4.5–88.4)</td>
<td>23.3 hrs (4.7–144.1)</td>
</tr>
<tr>
<td>Mean WRT, Short Category (≤10th percentile)</td>
<td>10.1 hrs (s.d. 0.9)</td>
<td>5.9 hrs (s.d. 0.7)</td>
</tr>
<tr>
<td>Mean WRT, Medium Category (11th–89th percentile)</td>
<td>33.4 hrs (s.d. 9.7)</td>
<td>18.5 hrs (s.d. 7.8)</td>
</tr>
<tr>
<td>Mean WRT, Long Category (≥90th percentile)</td>
<td>74.4 hrs (s.d. 11.8)</td>
<td>60.4 hrs (s.d. 32.6)</td>
</tr>
</tbody>
</table>

* The period of study changed between the two analyses because a different hydraulic model was used to derive WRT estimates in Utility 1 from 1996–1998 vs. 1999–2003 and in the present analysis there was not sufficient sample size to run a separate analysis for the time period from 1996–1998.

** # of ED visits analyzed includes all visits for GI illness (cases) + all non-injury non-GI illness visits (comparison group)
Table 2

Results of logistic regression model estimating the odds of emergency department visits for gastrointestinal illness as a function of Euclidian distance between households and the nearest node in the drinking water distribution system. Models controlled for Medicaid status, median household income at the census block level, percent minority at the census block level, patient age, Euclidean distance between hospital where ED visit occurred and geocoded patient residence, indicator variables for hospital, year, day-of-week, and product terms for age*hospital, age* distance to hospital, age*Medicaid status.

<table>
<thead>
<tr>
<th>Distance to Nearest Node</th>
<th>100–500 ft vs. &lt;100 ft</th>
<th>&gt;500 ft vs. &lt;100 ft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility 1</td>
<td>N=219,093</td>
<td>OR=1.02 (0.98–1.06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR=1.09* (1.03–1.15)</td>
</tr>
<tr>
<td>Utility 2</td>
<td>N=473,211</td>
<td>OR=0.98 (0.95–1.01)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OR=0.96 (0.92–1.00)</td>
</tr>
</tbody>
</table>