Detangling Seasonal Relationships of Fecal Contamination Sources and Correlates with Indicators in Michigan Watersheds

Amanda M. Wilson,[a] Sherry L. Martin,[b] (ORCID) Marc P. Verhougstraete,[a] Anthony D. Kendall,[b] Amity G. Zimmer-Faust,[c] Joan B. Rose,[d] Melanie L. Bell,[e] David W. Hyndman[b,f]

[a]Department of Community, Environment & Policy, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona, USA
[b]Department of Earth and Environmental Sciences, College of Natural Science, Michigan State University, East Lansing, Michigan, USA
[c]Southern California Coastal Water Research Project Authority, Costa Mesa, California, USA
[d]Department of Fisheries and Wildlife, College of Agriculture and Natural Resources, Michigan State University, East Lansing, Michigan, USA
[e]Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona, USA
[f]Department of Geosciences, School of Natural Sciences and Mathematics, University of Texas, Richardson, Texas, USA

ABSTRACT Despite the widely acknowledged public health impacts of surface water fecal contamination, there is limited understanding of seasonal effects on (i) fate and transport processes and (ii) the mechanisms by which they contribute to water quality impairment. Quantifying relationships between land use, chemical parameters, and fecal bacterial concentrations in watersheds can help guide the monitoring and control of microbial water quality and explain seasonal differences. The goals of this study were to (i) identify seasonal differences in *Escherichia coli* and *Bacteroides thetaiotaomicron* concentrations, (ii) evaluate environmental drivers influencing microbial contamination during baseflow, snowmelt, and summer rain seasons, and (iii) relate seasonal changes in *B. thetaiotaomicron* to anticipated gastrointestinal infection risks. Water chemistry data collected during three hydroclimatic seasons from 64 Michigan watersheds were analyzed using seasonal linear regression models with candidate variables including crop and land use proportions, prior precipitation, chemical parameters, and variables related to both wastewater treatment and septic usage. Adaptive least absolute shrinkage and selection operator (LASSO) linear regression with bootstrapping was used to select explanatory variables and estimate coefficients. Regardless of season, wastewater treatment plant and septic system usage were consistently selected in all primary models for *B. thetaiotaomicron* and *E. coli*. Chemistry and precipitation-related variable selection depended upon season and organism. These results suggest a link between human pollution (e.g., septic systems) and microbial water quality that is dependent on flow regime.

IMPORTANCE In this study, a data set of 64 Michigan watersheds was utilized to gain insights into fecal contamination sources, drivers, and chemical correlates across seasons for general *E. coli* and human-specific fecal indicators. Results reaffirmed a link between human-specific sources (e.g., septic systems) and microbial water quality. While the importance of human sources of fecal contamination and fate and transport variables (e.g., precipitation) remain important across seasons, this study provides evidence that fate and transport mechanisms vary with seasonal hydrologic condition and microorganism source. This study contributes to a body of research that informs prioritization of fecal contamination source control and surveillance strategy development to reduce the public health burden of surface water fecal contamination.

KEYWORDS *B. thetaiotaomicron*, *E. coli*, watershed, water quality, fecal indicator, septic systems

The financial and public health impacts of fresh water fecal contamination have long been recognized: fecal contamination of freshwater is a public health concern with potential to degrade the quality of irrigation water for food crops (1, 2), groundwater used for drinking water (3), and surface waters that have multiple beneficial uses (4), yielding significant disease and financial burdens. Direct exposure to contaminated surface waters, such as those through recreational activities, can also lead to negative health outcomes (5, 6). For example, an epidemiological study at beaches in Brazil found direct correlations between *Enterococcus* concentrations in water and gastrointestinal illness rates in swimmers (7). Freshwater beach epidemiological studies have correlated *Escherichia coli* concentrations in water with gastrointestinal illnesses in swimmers (8, 9).

Two commonly utilized methods for investigating fecal contamination sources in surface waters are (i) indicator organisms and (ii) microbial source tracking (MST) of host-specific gene markers. Indicator organisms (e.g., *E. coli*) are not specific to a particular animal species. MST methods utilize PCR-based methods to quantify specific gene sequences associated with different sources (10–12). One example of MST is the use of single-copy homologues encoding *Bacteroides thetaiotaomicron* α-1-6 mannanases found in high percentages of human fecal samples and indicative of human fecal contamination when detected in water (13, 14).

Some transport factors influencing fecal contamination in water are well recognized, such as runoff from land into surface waters (15, 16). However, the specific mechanisms of fecal contamination fate and transport and corresponding public health burdens are not well understood. Improving surveillance protocols for prioritizing fecal contamination source control will help to implement evidence-based public health interventions.

Contaminant occurrences and concentrations depend on myriad factors, including contamination sources, hydrologic conditions (17), physical properties of the water, weather (18–20), population density, land use characteristics, and human behavior (21). In temperate regions, some of these parameters are seasonally driven, resulting in water quality differences within any given year. As mentioned previously, precipitation and hydrologic conditions are well-known drivers of fecal contamination. A previous Great Lakes study found inconsistent relationships between fecal indicator bacteria and human-specific markers under low- and high-flow conditions, suggesting that hydrologic conditions play a role in water contamination (20). However, parsing out individual fates and transport mechanisms for specific fecal contaminants can be challenging in complex systems that receive multiple sources of water and pollution.

Relating *E. coli* and *B. thetaiotaomicron* fate and transport mechanisms with seasonal environmental changes would improve the understanding of risks posed by fecal contamination in surface waters as different recreational activities vary seasonally. Therefore, this study sought to (i) identify seasonal differences in *E. coli* and *B. thetaiotaomicron* concentrations in surface waters, (ii) determine environmental drivers of microbial contamination in baseflow, snowmelt, and summer rain seasons, and (iii) use quantitative microbial risk assessment (QMRA) across seasons to address exposures from recreational activities and associated gastrointestinal illnesses. Quantifying relationships between fecal contamination, land use or environmental parameters, and human health can provide a basis to improve watershed management policies and microbial risk mitigation.

RESULTS

In the 64 watersheds studied across a large (88,000-km^2) spatial scale, there were significant seasonal differences in $\log_{10}$ concentrations of *B. thetaiotaomicron* ($P < 0.001$), a human fecal source tracking marker (Fig. 1; see also Table S1 in the supplemental material). Mean $\log_{10}$ concentrations of *B. thetaiotaomicron* varied over multiple orders of magnitude. Concentrations were highest during baseflow (5.15 $\log_{10}$ cell equivalents [CE]/100 mL), followed by the snowmelt (3.84 $\log_{10}$ CE/100 mL) and summer rain (3.07 $\log_{10}$ CE/100 mL) seasons (Fig. 1). This trend was routinely measured across watersheds (Fig. 1). In contrast, mean $\log_{10}$ *E. coli* concentrations, while statistically significantly different ($P < 0.002$) across
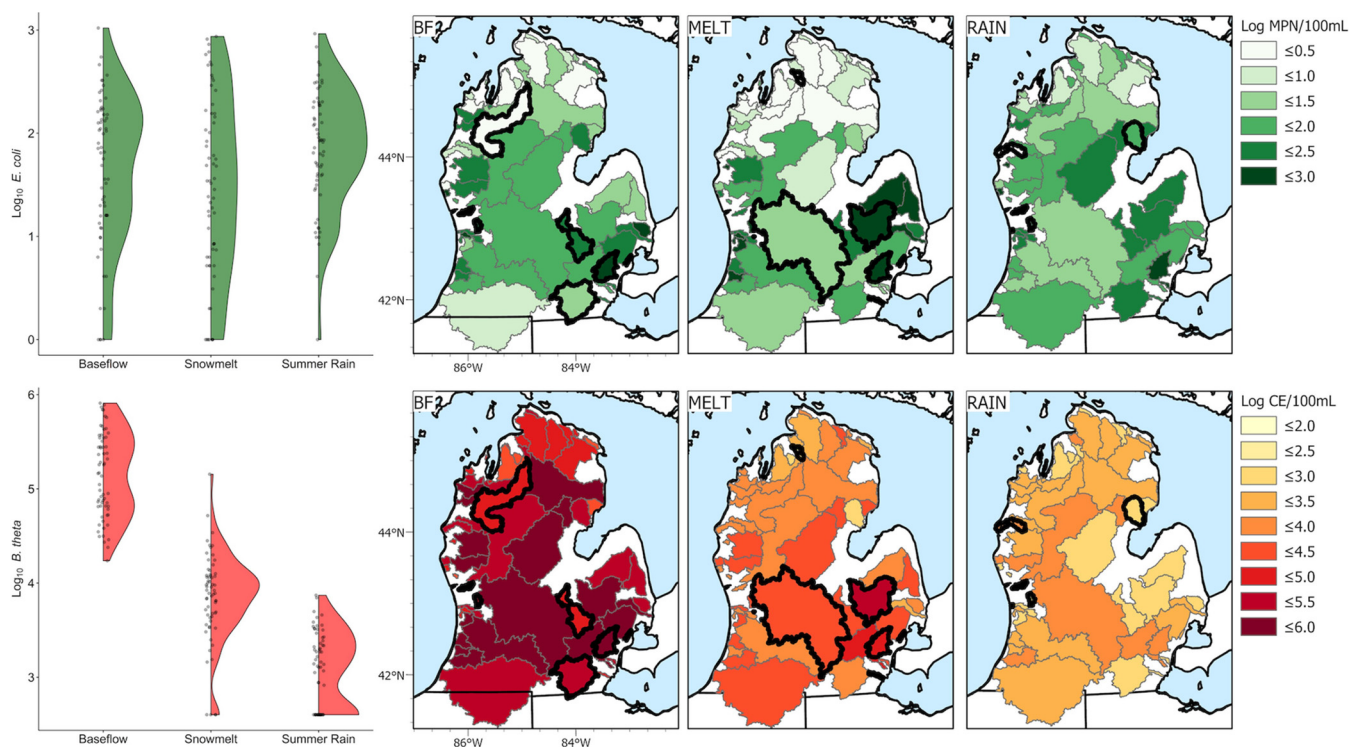
FIG 1 Distributions of *B. thetaiotaomicron* (*B. theta*) and *E. coli* log$_{10}$ concentrations across seasons in chronological order of measurement collection (baseflow [BF], snowmelt, and summer rain). Each watershed sampled in this study is shown with corresponding color and gray border. Bolded watersheds were identified as influential watersheds in the sensitivity analysis. Baseflow measurements were collected from 1 to 13 October 2010. Snowmelt and summer rain measurements were collected from 5 to 23 February 2011 and from 1 to 28 June 2011, respectively.

seasons, were within the same order of magnitude (~1 log$_{10}$; baseflow, 1.81 log$_{10}$; snowmelt, 1.53 log$_{10}$; summer rain, 1.93 log$_{10}$), with no strong spatial pattern across seasons (Fig. 1). Most precipitation and chemical parameters were significantly different across seasons (Table S1).

Notable least absolute shrinkage and selection operator (LASSO)-selected variables and their seasonal trends. Total dissolved nitrogen was selected in multiple primary and sensitivity models with statistically significant estimated beta coefficients (Fig. 2). While primarily selected for *E. coli*, one model also selected total dissolved nitrogen for *B. thetaiotaomicron* (summer rain, sensitivity). Based on these beta coefficients, the *E. coli* baseflow model found that an increase of 1 μg/L of total dissolved nitrogen was associated with a 36% $[(e^{3.1 \times 10^{-1}} - 1) \times 100\%]$ increase in *E. coli* concentrations (most probable number [MPN]/100 mL). In the snowmelt model, an increase of 1 μg/L in total dissolved nitrogen was related to a 20% $[(e^{1.8 \times 10^{-1}} - 1) \times 100\%]$ increase in *E. coli* concentrations (MPN/ 100 mL) (Table S2).

Other consistently selected variables for all seasons related to *B. thetaiotaomicron* human-specific fecal contamination sources included WWTP- and septic-related variables (Fig. 2). Two of the human pollution-related variables (e.g., percentage of the population on septic, percentage of the population on WWTPs, number of septic systems in watersheds, or population density) were found to be important drivers of microbial water quality in all primary models (Fig. 2). Specifically, the percentage of the population on septic systems was selected in 8 of the 12 primary and sensitivity analysis models, and the percentage of the population on WWTPs was selected in 4 of 12 models (Fig. 2; Table S2). WWTP variables (e.g., percent population on WWTPs and number of people on WWTPs) and number of septic tanks in watersheds consistently had positive relationships with log-transformed *E. coli* and *B. thetaiotaomicron* concentrations, in both primary and sensitivity analysis models (Fig. 2; Table S2). Despite having been selected in all models, only the baseflow sensitivity model for *B. thetaiotaomicron* had
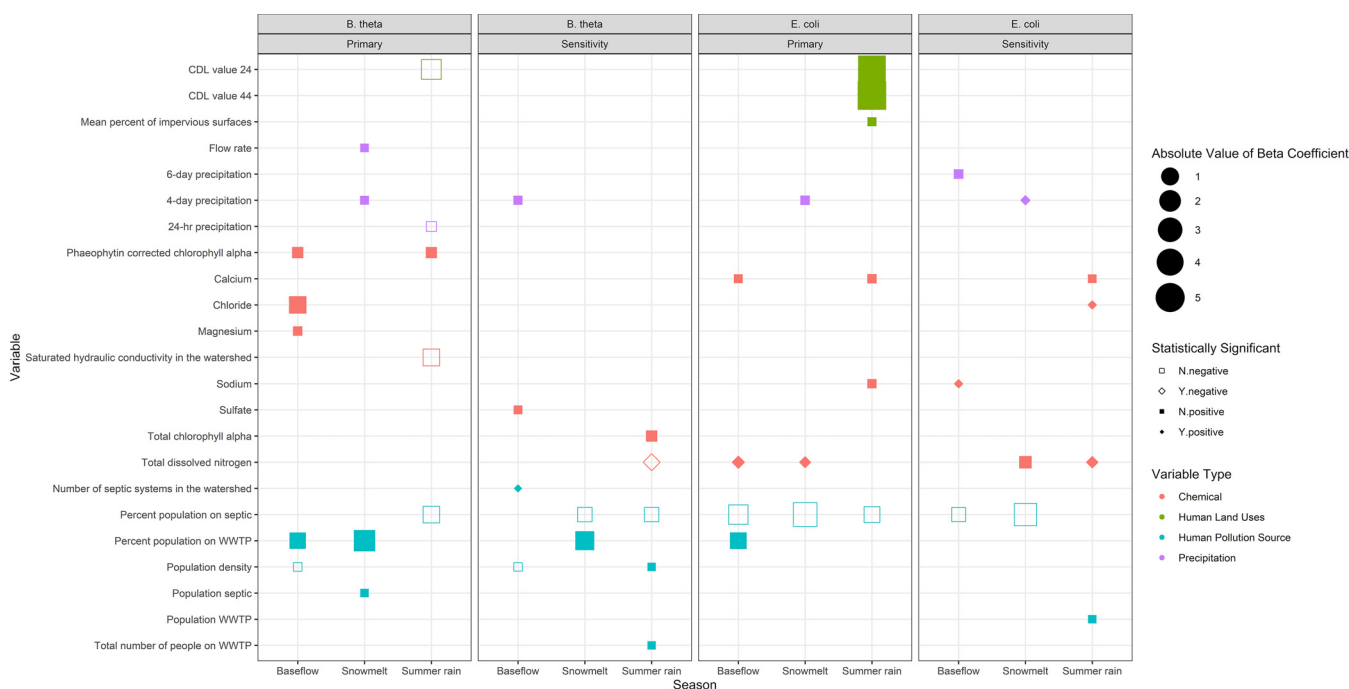
FIG 2 Selected and statistically significant explanatory variables for primary and sensitivity *E. coli* and *B. thetaiotaomicron* models. Variable names are described in more detail in Table S2. The total numbers of observations (i.e., watersheds) used in the primary models for *E. coli* and *B. thetaiotaomicron* per season were 56, 46, and 60 for the baseflow, snowmelt, and summer rain models, respectively. In sensitivity analysis *E. coli* models, 2, 3, and 4 influential watersheds were removed from the baseflow, snowmelt, and summer rain data sets, respectively. In the sensitivity analysis *B. thetaiotaomicron* models, 5, 4, and 2 influential watersheds were removed from baseflow, snowmelt, and summer rain data sets, respectively. Missing data are described in the supplemental material. N.negative, not statistically significant negative relationship; Y.negative, statistically significant negative relationship; N.positive, not statistically significant positive relationship; Y.positive, statistically significant positive relationship.

a statistically significant beta coefficient for one of these human pollution source variables.

Relating seasonal changes in *B. thetaiotaomicron* to anticipated risks. Temporal changes in *B. thetaiotaomicron* concentrations were related to anticipated infection risks using a QMRA approach to illustrate how MST, hydrologic methods, and QMRA can be collectively used. Norovirus, a human pathogen associated with fecal contamination, was used as the pathogen of interest because it is waterborne and has existing dose-response curves (22). Additional modeling details are described in Materials and Methods and in the supplemental material. Briefly, the concentrations of *B. thetaiotaomicron* measured in this study were related to previously reported *B. thetaiotaomicron* and norovirus concentrations in septic systems. The ratio of *B. thetaiotaomicron* to norovirus was then used to estimate norovirus concentrations in the watershed samples, assuming that norovirus was always present in the septic source (i.e., fluctuations in norovirus and *B. thetaiotaomicron* concentrations were not occurring and therefore the ratio was constant). Relative to mean infection risks during the baseflow season, mean infection risks for an hour of recreational activity with water ingestion in the summer rain and snowmelt seasons were decreased by 95% and 79%. A sensitivity analysis of this risk-based approach identified calculated norovirus concentration based on *B. thetaiotaomicron* concentrations in surface water as the most influential parameter on estimation of infection risks for all three seasons ($\rho = 0.9996$), followed by the pathogen concentrations in the source septic systems ($\rho = 0.79$) (Fig. S2).

DISCUSSION

Mechanistic explanations for model selection of WWTP and septic variables. The positive relationships observed between both WWTPs and the total number of septic tanks with log-transformed *E. coli* and *B. thetaiotaomicron* indicate that contamination in the watersheds evaluated in this study may be primarily driven by human
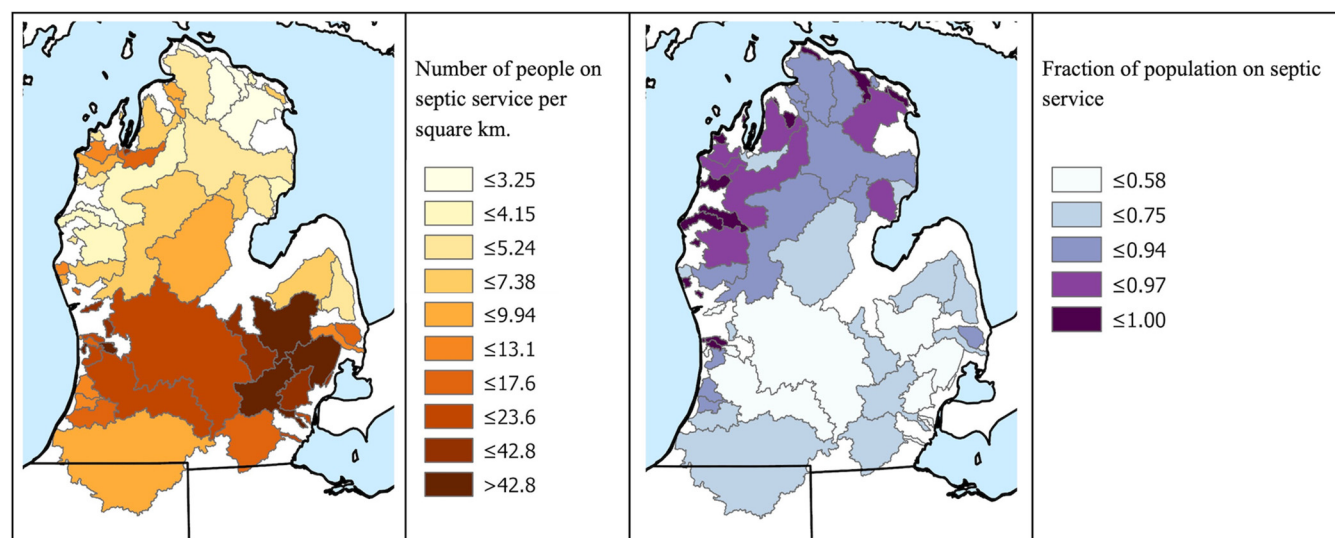
FIG 3 Spatial distribution of fraction of the population on septic service and number of people on septic service per square kilometer. Data are shown in quantiles, where 20% of the data are shown in each category. Each watershed sampled in this study is shown with corresponding color and gray border.

fecal sources. Identification of septic systems and WWTPs as inputs of surface water fecal contaminations in this study is consistent with existing literature (23), and there is prior evidence of WWTP and septic system contributions to the watersheds evaluated (24–27). Either a variable related to WWTPs (e.g., percentage of the population on WWTPs and number of people on WWTPs) or total number of septic systems was selected regardless of hydrologic condition, implying that inputs of human fecal contamination likely occur year-round.

Counterintuitively, negative relationships between the percentage of the population on septic and log-transformed *E. coli* and *B. thetaiotaomicron* concentrations were observed. However, these relationships were likely confounded by structural differences in watersheds across the state, leading to covariation between underlying drivers of fecal contamination. In Michigan's northern Lower Peninsula, most of the population that is served by septic systems (Fig. 3) lives within largely forested watersheds with limited agricultural activity. Here, a negative relationship between percentage of the population on septic and *E. coli* was observed that was likely driven in part by the low population density (Fig. 3). This region had relatively low *E. coli* concentrations overall, especially during baseflow and snowmelt seasons (Fig. 1). A similar trend was present for *B. thetaiotaomicron*, where concentrations were greater in watersheds in the center of the state than in the northern area of the state across seasons (Fig. 1).

In contrast, the number of septic systems in the watersheds was positively correlated with *B. thetaiotaomicron* concentrations during baseflow, with a significant model beta coefficient. This may be explained by *B. thetaiotaomicron* loading, predominately due to septic system failures. While widespread data on septic system failures are unavailable, they are likely driven by factors such as system age, usage, type, and maintenance along with soil characteristics and hydrology. Though there are spatial differences in these factors across watersheds, septic failures are a relatively rare event and should thus be viewed probabilistically. Specifically, a greater number of septic systems upstream of a sampling location means a greater probability of one or more system failures. As a result, in this study, total number of septic systems in the watershed may be a more appropriate surrogate for potential septic impact on water quality than percentage of the population on septic.

Model selection was also impacted by the methods utilized. Adaptive LASSO, like all regression shrinkage models, can underfit parameters, resulting in low predictive power (28). This statistical limitation could explain, in part, the low adjusted R^2 values and relatively small beta parameter estimates in this study. However, adaptive LASSO

is an appropriate approach in the case where there are many candidate variables for few outcome variables, as was the case in this study. Additionally, while collinearity among candidate explanatory variables was a challenge, ridge regression was used to shrink the weights for colinear variables (29, 30).

Mechanisms for similarities between models of *E. coli* and *B. thetaiotaomicron*. Some variables were shared by *E. coli* and *B. thetaiotaomicron* models for the same season, providing evidence of relationships between general fecal contamination and human-associated contamination. This also suggests that *E. coli* and *B. thetaiotaomicron* may share a source and/or follow similar fates and transport mechanisms when entering the environment. Under baseflow conditions, the percentage of the population on WWTPs was selected for both *E. coli* and *B. thetaiotaomicron* primary models (Fig. 2; Table S2). This explanatory variable had a positive relationship with *E. coli* and *B. thetaiotaomicron* concentrations, indicating the importance of direct human contributions of fecal contamination as a source for both organisms. Under snowmelt conditions, 4-day precipitation was selected in both primary models, where it had a positive relationship with *E. coli* and *B. thetaiotaomicron* concentrations (Table S2), indicating that there may be shared transport mechanisms for these organisms.

Mechanisms for differences across seasons. The consistent selection of a certain category of variable across seasons indicates its year-round relevance as a source or as fate and transport mechanisms. Precipitation parameters were selected across seasons in primary models (Fig. 2). However, the direction of their relationship with fecal indicators was inconsistent, implying that while it may be mechanistically important year-round, the role of precipitation in fate and transport may change. For example, in the snowmelt season, rainfall had a positive relationship with *E. coli* and *B. thetaiotaomicron* concentrations (Table S2), indicating enhanced delivery. However, in the summer rain season, rainfall had a negative relationship with log-transformed *B. thetaiotaomicron* concentrations (Table S2), indicating dilution effects. The positive relationship between precipitation and fecal indicators was expected, especially in conjunction with the model selection of total dissolved nitrogen: increased overland flow runoff due to precipitation in agricultural areas is expected to increase nitrogen and *E. coli* loading in watersheds (31). However, the negative summer relationship between precipitation and *E. coli* has not been reported, to our knowledge. This could be indicative of a dilution effect caused by increased overland flow rates in comparison to subsurface flow, where septic tank effluent is present. Increased surface flows may also limit the environmental persistence and regrowth of these organisms.

Elucidating the relative importance of fecal contamination sources and fate and transport mechanisms over time and across locations has the potential to reduce recreational risks (32), improve water quality surveillance efforts (33), and help prioritize fecal contamination source control (18, 34). Concentrations of the human-specific *B. thetaiotaomicron* were the most influential parameter related to infection risk. Thus, parameters that explain or signal seasonal changes in *B. thetaiotaomicron* and human fecal contamination concentrations and their relationship to pathogens would be useful for anticipating seasonal changes in risk. To better estimate absolute risks, future work could address and improve upon the assumptions within our simple QMRA model framework. For example, capturing the growth and decay dynamics of *E. coli* and *B. thetaiotaomicron* after their release from sources would allow for more accurate analysis of seasonal relationships and subsequent risk estimates. Additional data that help improve risk assessments include seasonal changes in the number of people who engage in exposure-related activities, parameters that relate differences in the transport of pathogens versus indicator organisms from sources to sampling sites, and mathematical explanations for anticipated mechanistic changes in septage inputs to surface waters as a function of season.

Conclusions. This study provides further evidence that WWTPs and septic systems are important, year-long sources of human fecal contamination within the evaluated watersheds. Furthermore, we show that the relationship between contamination and other explanatory variables is seasonally dependent. However, there are still uncertainties regarding

the magnitude of these sources relative to other fecal contamination inputs, which is likely geographically and temporally specific. Future research is needed to quantify the impact of human-specific sources on surface water fecal contamination, accounting for geographical and temporal effects. In particular, discerning impacts from septic systems was challenging due to confounding factors and limited reporting regarding septic system integrity and usage.

The analyses from this study also demonstrate the year-long importance of precipitation in fate and transport mechanisms of fecal contamination, but inconsistencies in the direction of the relationship between precipitation and fecal indicator concentrations indicate that the role of precipitation in these mechanisms is likely seasonally driven. Other parameters, including total dissolved nitrogen, appear to be important correlates of *E. coli* concentrations during baseflow and snowmelt seasons and are commonly related to other watershed inputs such as agriculture. Further research is needed to understand causal mechanisms behind the relationships identified here and to quantify relationships with more statistical confidence, using larger sample sizes or other statistical methods not as limited by underfitting and low predictive power. Additionally, more data per season would provide more confidence in seasonal differences (or lack thereof) elucidated in this study.

MATERIALS AND METHODS

Study region and sample size. An observational study was conducted in 64 watersheds across the state of Michigan across three hydrologic seasons: baseflow parameters were measured in fall (October 2010), snowmelt/spring thaw (March 2011), and postplanting summer rain (June 2011) (23). These intervals were selected to provide a broad range of hydrologic conditions across the contributing watersheds and within sampled streams. Details regarding sample collection timing and differences in seasonal streamflow are provided in the supplemental material.

At each site, streamflow was measured and water samples were collected as described by Verhougstraete et al. (23). Briefly, grab samples were collected from the center of each stream, at the upstream side of each access location. Biological samples were filtered and preserved within 24 h of collection. Samples for other water chemistry parameters were filtered in the field and chilled or frozen as appropriate. Laboratory analyses were performed within standard holding times and summary (23). Stream flows were either measured directly via appropriate methods or taken from 15-min data at colocated U.S. Geological Survey (USGS) streamflow gauges.

Variables. At each sampling location, physical and chemical parameters were measured during each sampling event. These included ammonium, calcium, chloride, dissolved oxygen, magnesium, nitrate/nitrite, total chlorophyll *a*, pheophytin-corrected chlorophyll *a*, pH, specific conductance, total dissolved nitrogen and phosphorus, sodium, potassium, dissolved organic carbon, and sulfate. Analytical methods have been described in detail by Verhougstraete et al. (23). *E. coli* was quantified using cultivation methods (IDEXX; Colilert), and the *B. thetaiotaomicron* genetic marker was measured using quantitative PCR (qPCR) (Roche; LightCycler 480 instrument) and previously published primer and probe sequences (14). Water temperature and streamflow were also measured during each sampling.

To quantify watershed precipitation prior to each sampling and static aspects of the landscape, watersheds were computed for each sampling location. For this, we used the 1/3 arc-second digital elevation model (DEM) from the National Elevation Data (NED) set, using a standard workflow in ArcMap (version 10.2): (i) the "fill" tool was run on the raw DEM, (ii) the "flowdirection" algorithm was run (which routes flow in one of the cardinal and subcardinal directions only—a D8 algorithm), (iii) the "flowaccumulation" tool was then run to make sure that sampling locations accurately fell within the DEM-generated flow network, and (iv) "watershed" was run to compute watershed areas for each sampling location.

These watersheds were then used to extract static characteristics from geographic information system (GIS) data sets that might reflect the influence of human or natural factors on *B. thetaiotaomicron* and *E. coli* concentrations. Land use and land cover data for each watershed were obtained from the USDA's Cropland Data Layer (CDL) and the USGS's National Land Cover Database (NLCD) 2006 values. Watershed variables associated with WWTP use, such as percentage of the population using septic tanks, number of septic systems, spatial density of septic systems, number of people on WWTPs, percentage of the population on WWTPs, and population density, were estimated using U.S. Census Bureau data as described by Luscz et al. (35).

Finally, we calculated nine spatially averaged precipitation variables within each watershed to account for precipitation totals (in millimeters) prior to sample collection in each watershed over different periods of time (6, 12, 18, and 24 h and 2, 3, 4, 6, and 8 days). Precipitation data were sourced from the North American Land Data Assimilation System (NLDAS) version 2, forcing data set A.

Statistical methods. Descriptive statistics were calculated for *B. thetaiotaomicron* and *E. coli* and for each sampling event: baseflow, snowmelt, and summer rain. Mixed-model analysis of variance was used to test for differences across the three seasons. Repeated measures in the same location were considered using an unstructured covariance matrix. Watersheds with missing values for candidate explanatory variables were not used in the primary or sensitivity analysis models.

The primary analysis was linear regression on log-transformed *B. thetaiotaomicron* and *E. coli* concentrations with explanatory variables selected with adaptive least absolute shrinkage and selection operator (LASSO) methods (30). Multiple statistical approaches have been used to explore environmental-microbial interactions, but to our knowledge, few have applied LASSO regression methods (36). One of the strengths of shrinkage methods includes their ability to address cases in which the number of potential explanatory variables is much larger than the number of observations (28), as is the case in this and many other stream health studies. Therefore, adaptive LASSO was used to select explanatory variables for *E. coli* and *B. thetaiotaomicron* models for all 3 seasons (6 primary models).

Ridge regression, which shrinks coefficients of correlated explanatory variables "toward each other" (29), was used to weight each candidate explanatory variable in adaptive LASSO. This is recommended for data sets in which candidate explanatory variables may be colinear, as in this case (30). Adaptive LASSO was restricted to select a number of explanatory variables that was equal to or less than the number of watersheds divided by 10 (*n*/10). For regression analysis, the outcome variables, *E. coli* and *B. thetaiotaomicron* concentrations, were log transformed. Bootstrapping methods were used to obtain 95% confidence intervals (CI) for regression parameters (beta parameter estimates) (37). Ten thousand subsets from the total data set were created using sampling with replacement, and adaptive LASSO was used to select variables from each of these subsets. If a variable was selected in more than 20% of the subset models, this variable was selected overall, and a 95% CI was reported for the beta parameters estimated per subset model.

Inclusion of zero in the 95% CI indicates uncertainty as to whether the selected variable had a relationship with the dependent variable. Therefore, exclusion of zero in the 95% CI for the beta parameter estimates indicated statistical significance at the 0.05 level. To obtain adjusted coefficients of determination (R^2), linear regression was used to fit log-transformed *E. coli* or *B. thetaiotaomicron* concentrations to their respective selected variables. After explanatory variables were selected by adaptive LASSO, linear regression assumptions were evaluated (details in the supplemental material).

A sensitivity analysis was conducted for each primary model to evaluate the effect of influential watersheds on variable selection and on adjusted R^2 values. Influential watersheds were identified using Cook's D plots and were removed from the sensitivity analysis prior to performing adaptive LASSO regression. Selected variables and the adjusted R^2 values for the models were then compared to those from the primary analysis that included influential watersheds. All statistical analyses were conducted using SAS 9.4 software (SAS Institute Inc., Cary, NC).

Relative-risk estimation. QMRA can estimate a probability of a health outcome such as infection, given a particular dose (38), and it has been used in surface water quality and recreational contexts (39–41). Differences in *B. thetaiotaomicron* concentrations across seasons were related to estimated relative risks of norovirus due to ingestion of water during recreational activities using a QMRA approach for relating fecal markers to reference pathogens (41). This approach has been used in multiple studies relating microbial source tracking in surface water to health outcomes (39–42). First, the *B. thetaiotaomicron* concentration at the watershed sampling locations (M_{sample}) was related to *B. thetaiotaomicron* concentrations at septic sources (M_{source}) to account for anticipated dilution from the source. A ratio of *B. thetaiotaomicron* concentration to norovirus concentration at the source (P_{source}) was then used to estimate how much norovirus would be present at the same watershed sampling location. This assumes that norovirus is always present at the source. The fraction of norovirus genome copies anticipated to be infectious ($f_{infectious}$) was then accounted for and multiplied by the estimated rate of water ingestion during recreational activities (V), informed by the work of Dorevitch et al. (43), and duration of the exposure (d), assumed to be equal to 1 h, consistent with another recreational water risk assessment (44), to yield an estimated dose (D) (equation 1):

$$D = \frac{M_{sample}}{M_{source}} \times P_{source} \times f_{infectious} \times V \times d \qquad (1)$$

While this assumes that the dilution of *B. thetaiotaomicron* and norovirus from septic sources would be the same, this approach was not used to calculate accurate risk. Rather, changes in risks estimated for summer rain and snowmelt seasons relative to baseflow were calculated. Data needed for a risk assessment are described in Discussion. A Monte Carlo approach was used to account for variability and uncertainty in input parameters. Explanations and sources for these distributions can be found in Table S3 and other supplemental material. As a sensitivity analysis, Spearman correlation coefficients were calculated to quantify linear relationships between input parameters and estimated infection risks. These results can be found in the supplemental material.

Distributions were fit to *B. thetaiotaomicron* concentrations for each season using the *fitdistrplus* package in R (version 4.0.2) (45), with details in the supplemental material. The effect of methods for handling left-censored data on distribution fitting was also explored (supplemental material). The distribution for concentrations of *B. thetaiotaomicron* at the source was informed by the work of Srinivasan et al. (46), and the distribution of norovirus at the source was informed by the work of Murphy et al. (47). A range of the fraction of genome copies representative of viable virus was randomly sampled from a uniform distribution with a minimum of 0.001 and a maximum of 0.01. The volume of water ingested during recreational activities was informed by the work of Dorevitch et al. (43). To account for differences in volumes ingested due to capsizing versus noncapsizing, two distributions were randomly sampled, where the volumes for capsized activities was randomly sampled 5.4% of the time to be consistent with the frequency at which capsizing occurred for these activities, according to Dorevitch et al. (43). Minimums and maximums of two uniform distributions were informed by minimum and maximum

mean estimated ingested volumes (in milliliters per hour) for a variety of activities and separately for capsized versus noncapsized activity.

Dose was then related to risk of infection using an approximate beta-Poisson dose-response curve described by Van Abel et al. (22), where $\alpha = 0.104$ and $\beta = 32.3$ (equation 2). Mean risks were used to estimate relative risks.

$$P_{\text{infect}} = 1 - (1 + \frac{D}{\beta})^{-\alpha} \qquad (2)$$

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 0.4 MB.

REFERENCES

1. Rajwar A, Srivastava P, Sahgal M. 2016. Microbiology of fresh produce: route of contamination, detection methods, and remedy. Crit Rev Food Sci Nutr 56:2383–2390. https://doi.org/10.1080/10408398.2013.841119.
2. Pearce-Walker J, Bright KR, Canales RA, Wilson AM, Verhougstraete MP. 2020. Managing leafy green safety from adenovirus and enteroviruses in irrigation water. Agric Water Manag 240:106272. https://doi.org/10.1016/j.agwat.2020.106272.
3. Hynds PD, Thomas MK, Pintar KDM. 2014. Contamination of groundwater systems in the US and Canada by enteric pathogens, 1990–2013: a review and pooled-analysis. PLoS One 9:e93301. https://doi.org/10.1371/journal.pone.0093301.
4. DeFlorio-Barker S, Wing C, Jones RM, Dorevitch S. 2018. Estimate of incidence and cost of recreational waterborne illness on United States surface waters. Environ Health 17:3. https://doi.org/10.1186/s12940-017-0347-9.
5. Probert WS, Miller GM, Ledin KE. 2017. Contaminated stream water as source for Escherichia coli O157 illness in children. Emerg Infect Dis 23: 1216–1218. https://doi.org/10.3201/eid2307.170226.
6. Verhougstraete MP, Pogreba-Brown K, Reynolds KA, Lamparelli CC, Zanoli Sato MI, Wade TJ, Eisenberg JNS. 2020. A critical analysis of recreational water guidelines developed from temperate climate data and applied to the tropics. Water Res 170:115294. https://doi.org/10.1016/j.watres.2019.115294.
7. Lamparelli CC, Pogreba-Brown K, Verhougstraete M, Sato MIZ, de Castro Bruni A, Wade TJ, Eisenberg JNS. 2015. Are fecal indicator bacteria appropriate measures of recreational water risks in the tropics: a cohort study of beach goers in Brazil? Water Res 87:59–68. https://doi.org/10.1016/j.watres.2015.09.001.
8. Marion JW, Lee J, Lemeshow S, Buckley TJ. 2010. Association of gastrointestinal illness and recreational water exposure at an inland U.S. beach. Water Res 44:4796–4804. https://doi.org/10.1016/j.watres.2010.07.065.
9. Dufour AP. 1984. Health effects criteria for fresh recreational waters. https://agris.fao.org/agris-search/search.do?recordID=US201300708833. Accessed 8 December 2021.
10. Ponce-Terashima R, Koskey AM, Reis MG, McLellan SL, Blanton RE. 2014. Sources and distribution of surface water fecal contamination and prevalence of schistosomiasis in a Brazilian village. PLoS Negl Trop Dis 8:e3186. https://doi.org/10.1371/journal.pntd.0003186.
11. Kildare BJ, Leutenegger CM, McSwain BS, Bambic DG, Rajal VB, Wuertz S. 2007. 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach. Water Res 41:3701–3715. https://doi.org/10.1016/j.watres.2007.06.037.
12. Layton BA, Cao Y, Ebentier DL, Hanley K, Ballesté E, Brandão J, Byappanahalli M, Converse R, Farnleitner AH, Gentry-Shields J, Gidley ML, Gourmelon M, Lee CS, Lee J, Lozach S, Madi T, Meijer WG, Noble R, Peed

L, Reischer GH, Rodrigues R, Rose JB, Schriewer A, Sinigalliano C, Srinivasan S, Stewart J, Van De Werfhorst LC, Wang D, Whitman R, Wuertz S, Jay J, Holden PA, Boehm AB, Shanks O, Griffith JF. 2013. Performance of human fecal anaerobe-associated PCR-based assays in a multi-laboratory method evaluation study. Water Res 47:6897–6908. https://doi.org/10.1016/j.watres.2013.05.060.
13. Kreader CA. 1995. Design and evaluation of Bacteroides DNA probes for the specific detection of human fecal pollution. Appl Environ Microbiol 61:1171–1179. https://doi.org/10.1128/aem.61.4.1171-1179.1995.
14. Yampara-Iquise H, Zheng G, Jones JE, Carson CA. 2008. Use of a Bacteroides thetaiotaomicron-specific α-1-6, mannanase quantitative PCR to detect human faecal pollution in water. J Appl Microbiol 105:1686–1693. https://doi.org/10.1111/j.1365-2672.2008.03895.x.
15. Lee SW, Hwang SJ, Lee SB, Hwang HS, Sung HC. 2009. Landscape ecological approach to the relationships of land use patterns in watersheds to water quality characteristics. Landsc Urban Plan 92:80–89. https://doi.org/10.1016/j.landurbplan.2009.02.008.
16. Liang Z, He Z, Zhou X, Powell CA, Yang Y, He LM, Stoffella PJ. 2013. Impact of mixed land-use practices on the microbial water quality in a subtropical coastal watershed. Sci Total Environ 449:426–433. https://doi.org/10.1016/j.scitotenv.2013.01.087.
17. Nshimyimana JP, Martin SL, Flood M, Verhougstraete MP, Hyndman DW, Rose JB. 2018. Regional variations of bovine and porcine fecal pollution as a function of landscape, nutrient, and hydrological factors. J Environ Qual 47:1024–1032. https://doi.org/10.2134/jeq2017.11.0438.
18. Mattioli MC, Benedict KM, Murphy J, Kahler A, Kline KE, Longenberger A, Mitchell PK, Watkins S, Berger P, Shanks OC, Barrett CE, Barclay L, Hall AJ, Hill V, Weltman A. 2021. Identifying septic pollution exposure routes during a waterborne norovirus outbreak—a new application for human-associated microbial source tracking qPCR. J Microbiol Methods 180: 106091. https://doi.org/10.1016/j.mimet.2020.106091.
19. Tolouei S, Burnet J-B, Autixier L, Taghipour M, Bonsteel J, Duy SV, Sauvé S, Prévost M, Dorner S. 2019. Temporal variability of parasites, bacterial indicators, and wastewater micropollutants in a water resource recovery facility under various weather conditions. Water Res 148:446–458. https://doi.org/10.1016/j.watres.2018.10.068.
20. Dila DK, Corsi SR, Lenaker PL, Baldwin AK, Bootsma MJ, McLellan SL. 2018. Patterns of host-associated fecal indicators driven by hydrology, precipitation, and land use attributes in Great Lakes Watersheds. Environ Sci Technol 52:11500–11509. https://doi.org/10.1021/acs.est.8b01945.
21. Peed LA, Nietch CT, Kelty CA, Meckes M, Mooney T, Sivaganesan M, Shanks OC. 2011. Combining land use information and small stream sampling with PCR-based methods for better characterization of diffuse

sources of human fecal pollution. Environ Sci Technol 45:5652–5659. https://doi.org/10.1021/es2003167.

22. Van Abel N, Schoen ME, Kissel JC, Meschke JS. 2017. Comparison of risk predicted by multiple norovirus dose-response models and implications for quantitative microbial risk assessment. Risk Anal 37:245–264. https://doi.org/10.1111/risa.12616.

23. Verhougstraete MP, Martin SL, Kendall AD, Hyndman DW, Rose JB. 2015. Linking fecal bacteria in rivers to landscape, geochemical, and hydrologic factors and sources at the basin scale. Proc Natl Acad Sci U S A 112:10419–10424. https://doi.org/10.1073/pnas.1415836112.

24. Ahmed W, Neller R, Katouli M. 2005. Evidence of septic system failure determined by a bacterial biochemical fingerprinting method. J Appl Microbiol 98:910–920. https://doi.org/10.1111/j.1365-2672.2004.02522.x.

25. Harris PJ. 1995. Water quality impacts from on-site waste disposal systems to coastal areas through groundwater discharge. Geo 26:262–268. https://doi.org/10.1007/BF00770477.

26. Iverson G, Sanderford C, Humphrey CP, Etheridge JR, Kelley T. 2020. Fecal indicator bacteria transport from watersheds with differing wastewater technologies and septic system densities. Appl Sci 10:6525. https://doi.org/10.3390/app10186525.

27. Conn KE, Habteselassie MY, Denene Blackwood A, Noble RT. 2012. Microbial water quality before and after the repair of a failing onsite wastewater treatment system adjacent to coastal waters: coastal wastewater microbial fate. J Appl Microbiol 112:214–224. https://doi.org/10.1111/j.1365-2672.2011.05183.x.

28. Hernandez-Suarez JS, Nejadhashemi AP. 2018. A review of macroinvertebrate- and fish-based stream health modelling techniques. Ecohydrology 11:e2022. https://doi.org/10.1002/eco.2022.

29. Friedman J, Hastie T, Tibshirani R. 2010. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33:1–22.

30. Ou HZ. 2006. The adaptive lasso and its oracle properties. J Am Stat Assoc 101:1418–1429. https://doi.org/10.1198/016214506000000735.

31. Poor CJ, McDonnell JJ. 2007. The effects of land use on stream nitrate dynamics. J Hydrol 332:54–68. https://doi.org/10.1016/j.jhydrol.2006.06.022.

32. Russo GS, Eftim SE, Goldstone AE, Dufour AP, Nappier SP, Wade TJ. 2020. Evaluating health risks associated with exposure to ambient surface waters during recreational activities: a systematic review and meta-analysis. Water Res 176:115729. https://doi.org/10.1016/j.watres.2020.115729.

33. Lothrop N, Bright KR, Sexton J, Pearce-Walker J, Reynolds KA, Verhougstraete MP. 2018. Optimal strategies for monitoring irrigation water quality. Agric Water Manag 199:86–92. https://doi.org/10.1016/j.agwat.2017.12.018.

34. Ervin JS, Van De Werfhorst LC, Murray JLS, Holden PA. 2014. Microbial source tracking in a coastal California watershed reveals canines as controllable sources of fecal contamination. Environ Sci Technol 48:9043–9052. https://doi.org/10.1021/es502173s.

35. Luscz E, Kendall A, Hyndman D. 2015. High resolution spatially explicit nutrient source models for the Lower Peninsula of Michigan. J Gt Lakes Res 41:618–629. https://doi.org/10.1016/j.jglr.2015.02.004.

36. Tibshirani R. 1996. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Methodol 58:267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

37. SAS. 2015. SAS/STAT 14.1 user's guide. The GLMSELECT procedure. SAS Institute Inc, Cary, NC.

38. Haas CN, Rose JB, Gerba CP. 2014. Quantitative microbial risk assessment. John Wiley & Sons, New York, NY.

39. Brown KI, Graham KE, Soller JA, Boehm AB. 2017. Estimating the probability of illness due to swimming in recreational water with a mixture of human- and gull-associated microbial source tracking markers. Environ Sci Process Impacts 19:1528–1541. https://doi.org/10.1039/c7em00316a.

40. Boehm AB. 2019. Risk-based water quality thresholds for coliphages in surface waters: effect of temperature and contamination aging. Environ Sci Process Impacts 21:2031–2041. https://doi.org/10.1039/c9em00376b.

41. Brown KI, Graham KE, Boehm AB. 2017. Risk-based threshold of gull-associated fecal marker concentrations for recreational water. Environ Sci Technol Lett 4:44–48. https://doi.org/10.1021/acs.estlett.6b00473.

42. McLellan SL, Sauer EP, Corsi SR, Bootsma MJ, Boehm AB, Spencer SK, Borchardt MA. 2018. Sewage loading and microbial risk in urban waters of the Great Lakes. Elementa 6:46. https://doi.org/10.1525/elementa.301.

43. Dorevitch S, Panthi S, Huang Y, Li H, Michalek AM, Pratap P, Wroblewski M, Liu L, Scheff PA, Li A. 2011. Water ingestion during water recreation. Water Res 45:2020–2028. https://doi.org/10.1016/j.watres.2010.12.006.

44. Gitter A, Mena K, Wagner K, Boellstorff D, Borel K, Gregory L, Gentry T, Karthikeyan R. 2020. Human health risks associated with recreational waters: preliminary approach of integrating quantitative microbial risk assessment with microbial source tracking. Water Switz 12:327. https://doi.org/10.3390/w12020327.

45. Delignette-Muller ML, Dutang C. 2015. fitdistrplus: an R package for fitting distributions. J Stat Softw 64:1–34.

46. Srinivasan S, Aslan A, Xagoraraki I, Alocilja E, Rose JB. 2011. Escherichia coli, enterococci, and Bacteroides thetaiotaomicron qPCR signals through wastewater and septage treatment. Water Res 45:2561–2572. https://doi.org/10.1016/j.watres.2011.02.010.

47. Murphy HM, Thomas MK, Schmidt PJ, Medeiros DT, McFadyen S, Pintar KDM. 2016. Estimating the burden of acute gastrointestinal illness due to Giardia, Cryptosporidium, Campylobacter, E. coli O157 and norovirus associated with private wells and small water systems in Canada. Epidemiol Infect 144:1355–1370. https://doi.org/10.1017/S0950268815002071.