

Risk factors for *Cryptosporidium* contamination in Minnesota public supply wells

James F. Walsh^a, Deanna P. Scher^{id b,*}, Jane R. de Lambert^b and Anita C. Anderson^b

^a Minnesota Department of Health (retired), Saint Paul, MN 55164-0975, USA

^b Minnesota Department of Health, Saint Paul, MN 55164-0975, USA

*Corresponding author. E-mail: deanna.scher@state.mn.us

 DPS, 0000-0001-8150-0290

ABSTRACT

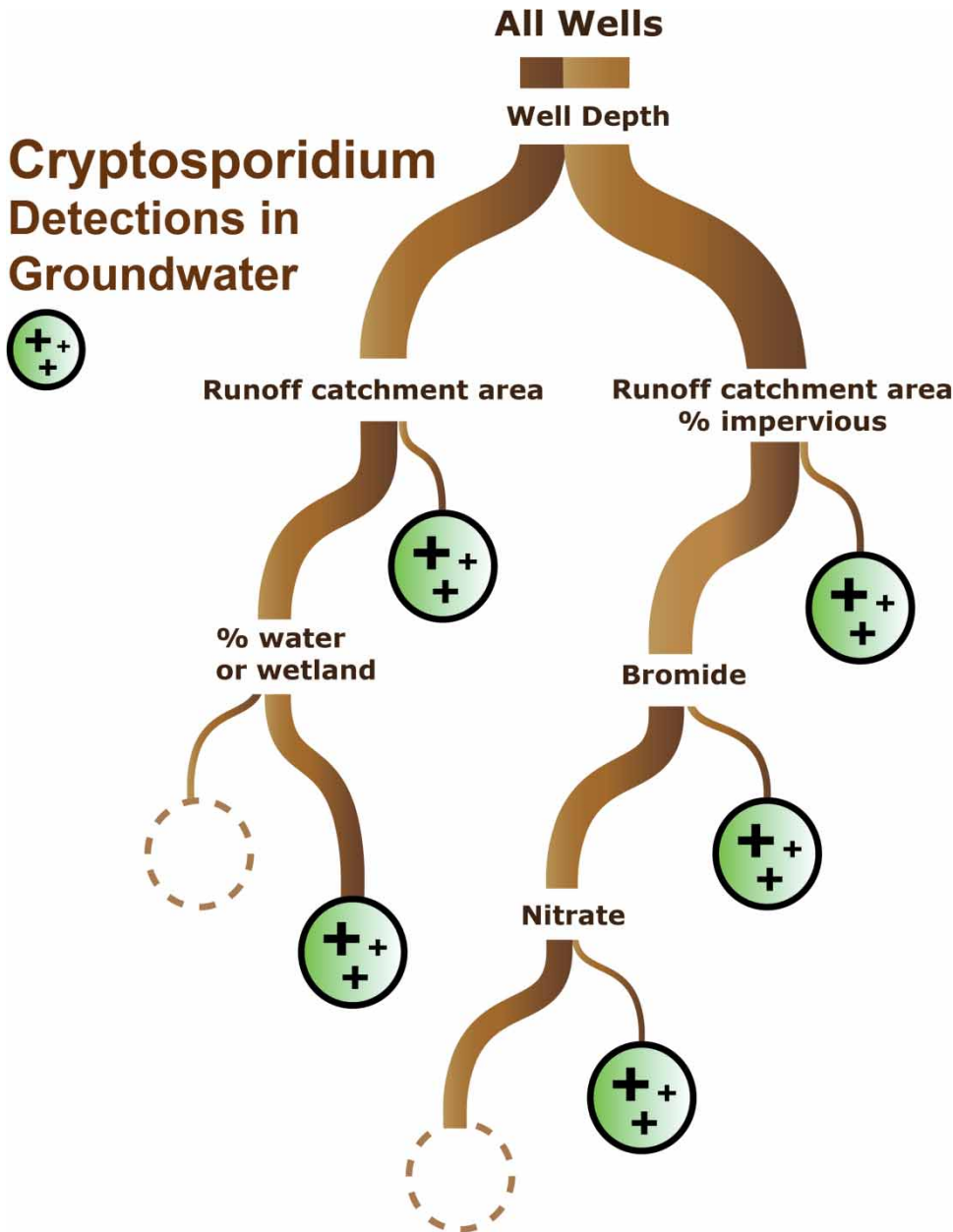
In a recent monitoring study of Minnesota's public supply wells, *Cryptosporidium* was commonly detected with 40% of the wells having at least one detection. Risk factors for *Cryptosporidium* occurrence in drinking water supply wells, beyond surface water influence, remain poorly understood. To address this gap, physical and chemical factors were assessed as potential predictors of *Cryptosporidium* occurrence in 135 public supply wells in Minnesota. Univariable analysis, regression techniques, and classification trees were used to analyze the data. Many variables were identified as significant risk factors in univariable analysis and several remained significant throughout the succeeding analysis techniques. These factors fell into general categories of well use and construction, aquifer characteristics, and connectedness to the land surface, well capture zones, and land use therein, existence of potential contaminant sources within 200-feet of the well, and variability in the chemical and isotopic parameters measured during the study. These risk categories, and the specific variables and threshold values we have identified, can help guide future research on factors influencing *Cryptosporidium* contamination of wells and can be used by environmental health programs to develop risk-based sampling plans and design interventions that reduce associated health risks.

Key words: *Cryptosporidium*, drinking water, groundwater contamination, microbial risk factors, public water supply wells, water monitoring

HIGHLIGHTS

- The influence of multiple factors on the presence of *Cryptosporidium* in public wells is examined.
- The human-caused and naturogenic risk factors, both well-understood and novel, are identified.
- The range of influential factors reflects varying pathways and sources of *Cryptosporidium*
- Findings can help drinking water managers assess risk and target monitoring efforts.

GRAPHICAL ABSTRACT



1. INTRODUCTION

The enteric parasite *Cryptosporidium* is the leading cause of waterborne disease among people in the United States (CDC 2023). Those with weakened immune systems are most at risk of developing serious and sometimes fatal illness (CDC 2023). Historically, *Cryptosporidium* has not been considered a contaminant of concern in groundwater sources of drinking water unless surface water is entering the wells (groundwater under direct influence, GWUDI). As a result, *Cryptosporidium* (and other protozoa) have rarely been included in groundwater monitoring studies (Hynds *et al.* 2014; Murphy *et al.* 2017), and U.S. regulations designed to prevent and monitor for *Cryptosporidium* contamination do not exist for groundwater, except GWUDI.

From 2014 to 2016, the Minnesota Department of Health (MDH) conducted a study to monitor for a diverse suite of microbiological contaminants in Minnesota, USA public water supply wells (Stokdyk *et al.* 2020). A total of 145 community and

noncommunity wells representing the major aquifer types across the state were repeatedly sampled over 1 or 2 years and tested for viral, bacterial, and protozoan pathogens and other indicators of fecal contamination using quantitative polymerase chain reaction (qPCR). Surprisingly, *Cryptosporidium* was found to be common in the public supply wells, with 40% of wells having at least one detection. Concentrations ranged from 0.05 to 246 gene copies per liter with a mean of 10.4 gene copies per liter. Gene sequencing to determine species was successful for 45 of the samples, with 41 identified as *Cryptosporidium parvum*, 2 as *Cryptosporidium andersoni*, and 2 as *Cryptosporidium hominis*, suggesting a range of possible sources, including humans, cattle, and other mammals. Of interest, *Cryptosporidium* detections were not found to be associated with surface water influence (Stokdyk *et al.* 2019). This finding is at odds with current risk and regulatory paradigms that assume limited occurrence of *Cryptosporidium* in groundwater. More recently, quantitative microbial risk assessment was used to quantify population-level health risk for nine viral, bacterial, and protozoan pathogens in the Minnesota study (Burch *et al.* 2022). *Cryptosporidium* was found to be a major contributor to health risk, making improved identification of wells susceptible to *Cryptosporidium* contamination a priority.

Risk factors for potential groundwater contamination by *Cryptosporidium*, beyond surface water influence, remain poorly defined (Chique *et al.* 2020). To our knowledge, no studies have systematically examined a wide range of hydrogeologic and anthropogenic factors that may influence *Cryptosporidium* occurrence in public or private wells across a diversity of hydrogeologic conditions. Most studies on potential risk factors have been targeted environmental investigations, prompted by human outbreaks of cryptosporidiosis, which are narrow in scope (Rose 1997; Watier-Grillot *et al.* 2022).

A better understanding of risk factors is especially imperative for *Cryptosporidium*, as routinely analyzed coliform indicator bacteria have been shown to be a poor predictor for the presence of enteric protozoa (Rose 1997; Stokdyk *et al.* 2020). Furthermore, once contamination occurs, the oocysts have a high survival rate (more than 24 months at 20°C) and a high resistance to disinfection (Betancourt & Rose 2004; Fradette *et al.* 2022). The primary objective of this work was to evaluate hydrogeologic and anthropogenic factors and routine general water chemistry parameters that may be associated with *Cryptosporidium* detections in public supply wells. The goal was to improve identification of risk factors for *Cryptosporidium* contamination, which would allow drinking water programs to target monitoring and interventions to reduce associated health risks.

2. METHODS

2.1. Study design and sample collection

The study was statewide in scope and included 145 wells from 123 public water systems (88 community wells and 57 noncommunity wells) in Minnesota (Figure 1). The wells were sampled every other month for one of two study phases (typically six samples per well per phase). The 89 wells in the first phase (May 2014–April 2015) were randomly selected from among non-disinfecting, year-round, community and nontransient noncommunity systems in Minnesota. For the second phase (May 2015–May 2016), 85 wells potentially vulnerable to contamination based on geologic setting, well construction, and past monitoring or groundwater flow modeling were prioritized, without regard to disinfection status, including 29 wells from Phase 1 (typically sampled 6 times per phase for a total of 12 samples). The two-phase selection process produced a sample of wells that generally represented the aquifers and geology of wells statewide, with a slight overemphasis on more vulnerable geologic settings.

Samples were collected by dead-end ultrafiltration (Smith & Hill 2009) using Hemodialyzer Rexeed-25s filters (Asahi Kasei Medical MT Corp., Oita, Japan). Groundwater volume sampled ranged from 140 to 1,783 L (mean, 728 L). All samples were collected prior to treatment (including disinfection) from sampling taps that were disinfected using a flame or alcohol wipes. Filters were shipped on ice and processed within 48 h.

Since pathogen occurrence was found to be sporadic in these study wells (Stokdyk *et al.* 2019, 2020), and additional studies have also found *Cryptosporidium* to occur intermittently in U.S. groundwater (Hancock *et al.* 1998), the outcome of interest in this analysis is any detection of *Cryptosporidium* in the well identified through the repeated sampling.

2.2. Study well characteristics

Well depth ranged from 20 to 630 feet, and age, casing length, and open interval length also varied (Table S1). Fifty-seven percent of the study wells drew from Quaternary glacial sand and gravel deposits; other aquifers included sandstone, fractured crystalline rock, carbonate rock, mixed sandstone, carbonate rocks, and shale (Table S2). The geologic sensitivity ratings of study wells, reflective of the estimated vertical time of travel for water moving from the land surface to the aquifer

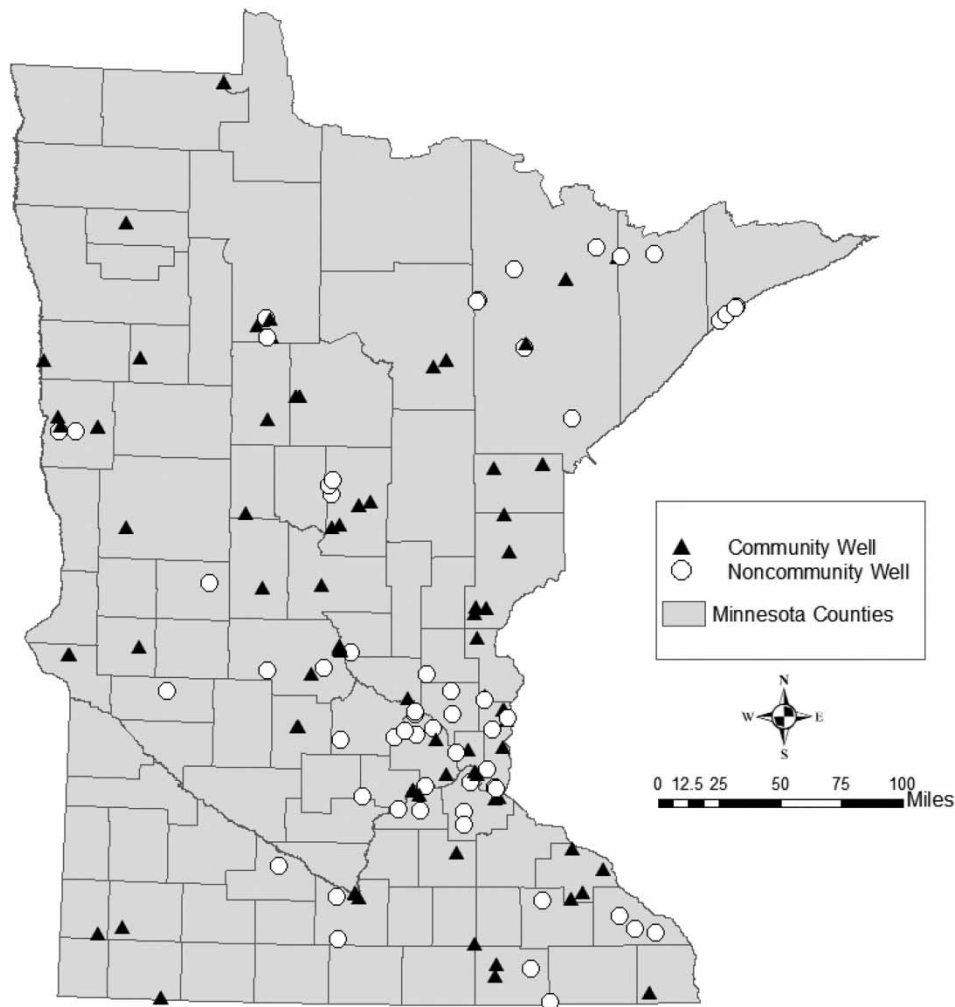


Figure 1 | Location of 145 sampled public water system wells.

in question, also varied. Study wells were considered representative of the ranges observed for all Minnesota public wells for these characteristics, with a slight bias towards wells with higher geologic sensitivity ratings.

2.3. Laboratory analysis

Analytical methods have been previously described (Stokdyk *et al.* 2016, 2019, 2020). Briefly, the samples were tested for *Cryptosporidium* by qPCR, targeting the 18S gene using a LightCycler 480 instrument (Roche Diagnostics, Mannheim, Germany) and following procedures and reaction conditions described in Stokdyk *et al.* (2016, 2019, 2020). qPCR was performed in duplicate, with the average of positive replicates reported.

In addition to the samples collected by ultrafiltration, samples were also collected for a variety of chemical and isotopic parameters indicative of wastewater influence and/or groundwater residence time and field measurements were taken for water temperature, specific conductance, pH, dissolved oxygen, and redox potential. Total coliform, *Escherichia coli*, total organic carbon, nitrate/nitrite nitrogen, ammonia, bromide, chloride, and boron were analyzed by the MDH Public Health Laboratory using standard methods (Table S3). Analyses of tritium and the stable isotopes of water (oxygen-18 and deuterium) were conducted at either the University of Waterloo Environmental Isotope Laboratory or Isotope Tracer Technologies. Deviation of stable isotope samples from the meteoric water line for central Minnesota (Landon *et al.* 2000) was determined using the line-conditioned excess approach (Landwehr & Coplen 2004).

2.4. Potential risk factors

A total of 79 hydrogeologic and landform/land use/land cover variables were selected for risk factor analysis based on their: (1) likelihood of affecting pathogen occurrence and transport at the land surface and in the subsurface, (2) function as indicators of groundwater residence time or human-impacted water quality, and (3) availability in Minnesota datasets. Several data sources were consulted to create the final set of potential predictors (Table S4). These variables were grouped into five general themes: Well use and construction; Aquifer characteristics, connectedness between aquifer and land surface; Well capture zone and land use within capture zone; Potential contaminant sources in the Inner Wellhead Management Zone (IWMZ); and Chemical and isotopic parameters (Table 1). Land-use variables within the well capture zones were derived from the National Land Cover Database and include various types of development (low-, medium-, and high-intensity

Table 1 | Hydrogeologic and anthropogenic factors ($n = 79$) assessed by theme^a

Theme	Potential predictive factors	
Well use and construction	Well type	Grout material
	Year drilled	Percent casing grouted
	Well depth	Percent grout saturated
	Depth cased	Annular space
	Casing diameter	Casing jointing method
	Casing material	Saturated casing value
	Drilling method	Discharge rate
	Grouted (yes/no)	
Aquifer characteristics, connectedness between aquifer and land surface	Land surface elevation	Near surface pollution sensitivity
	Depth to bedrock	Vertical hydraulic gradient (mean)
	Bedrock interface distance	Hydraulic conductivity
	Aquifer type	Aquifer thickness
	Aquifer porosity type	Static water level
	Groundwater age from tritium	Drawdown
	Karst or fractured	Surface water class
	Geologic sensitivity	Surface water subset
Well capture zone and land use within capture zone	L score	Primary groundwater class, unbiased
	Capture zone area	Development mostly agriculture (y/n), 1-yr TT
	Runoff catchment area	Percent open water or wetland, 1-yr TT
	Runoff catchment area, percent impervious	Percent low-intensity development, 10-yr TT
	Percent low-intensity development, 1-yr TT ^b	Percent medium-intensity development, 10-yr TT
	Percent medium-intensity development, 1-yr TT	Percent high-intensity development, 10-yr TT
	Percent high-intensity development, 1-yr TT	Percent row crop or pasture, 10-yr TT
	Percent row crop or pasture, 1-yr TT	Development mostly agriculture (y/n), 10-yr TT
Potential contaminant sources in the IWMZ		Percent open water or wetland, 10-yr TT
	Number (nbr) of pathogen sources	Nbr of storm sewer lines
	Nbr of drainfields	Dist. to nearest storm sewer line
	Distance (dist.) to nearest drainfield	Sewer type
	Nbr of septic/sewage systems	Sewer age
	Dist. to nearest septic/sewage system	Design flow
	Nbr of sewer lines	Waste treatment type
	Dist. to nearest sewer line	
Chemical and isotopic parameters	Variance from average precipitation	Ammonia CV
	Temporal variability	Boron CV
	Nitrate >1 mg/L in past 5 years	Total organic carbon CV
	Source total coliform detection ≤5 years	Specific conductance CV
	Distrib. total coliform detect ≤5 years	Temperature CV
	Bromide coefficient of variation (CV)	δ ² H CV
	Chloride CV	δ ¹⁸ O CV
	Chloride/bromide CV	pH CV
Nitrate CV	Dissolved oxygen CV	

^aSee Tables S5–S9 for variable definitions, descriptive statistics, and univariable analysis results. ^bTT=travel time.

residential, commercial, and industrial) as well as agricultural land uses. Potential pathogen sources within the IWMZ included septic systems (tanks and drainfields), sewer lines, and sewage lift stations. Chemical and isotopic variability was evaluated by assessing an average of six rounds of samples collected on a bimonthly basis over a 12-month period. Variable definitions are found in Tables S5–S9.

2.5. Statistical analysis

Analyses were limited to wells with at least three samples available ($n = 135$). Descriptive statistics were compiled for each potential risk factor in Table 1. Next, visual plots and univariable statistical tests looked for relationships between the potential predictors and *Cryptosporidium* detection. For continuous variables, the Mann-Whitney U test was selected, as the explanatory variables of interest typically showed skewed distributions. The Chi-square test was used for categorical variables. For categorical variables that contained groups with small sample sizes, categories were combined into larger groupings for testing when logical and feasible. The Cochran–Armitage trend test was applied to ordinal variables.

In addition to serving a prescreening role for multivariable modeling, univariable analysis results were considered independently for two reasons. First, several risk factors could not be included in multivariable modeling because they had a high degree of missing data which would have reduced the statistical power to evaluate other risk factors. Second, some variables had to be removed prior to or during multivariable modeling due to their collinearity or overall interrelatedness with other covariates.

Multivariable modeling was conducted to allow each potential risk factor to be evaluated for its independent effect and strength of association in the presence of other factors. Explanatory variables with >20% missing data were excluded. To reduce the number of potential predictors in Table 1 to a feasible size for multivariable modeling, only variables with a strength of association of $p \leq 0.2$ in univariable analysis were included. The study phase (Phases 1, 2, or both) was included as an independent variable. This factor not only accounted for potential differences in climatological conditions between the sampling periods, but also differences in the number of observations per well. Six samples were available for most wells sampled in Phase 1 or Phase 2, whereas wells sampled during both phases typically had 12 samples, which could affect (i.e., bias) the corresponding probability of pathogen detection for the well. Prior to building the regression model, Pearson correlation coefficients were assessed to check for collinearity between continuous variables. Inclusion/exclusion decisions were made when high correlation was found (>0.75).

A modified Poisson regression model with robust error variance was used to estimate the relative risk of *Cryptosporidium* detection using the generalized estimating equations-based method of Zou (2004). The model uses a log link function with a Poisson distribution. This approach was taken because odds ratios in logistic regression can overestimate relative risk when the outcome is common (McNutt *et al.* 2003) and in this case, there is a high prevalence of wells with a *Cryptosporidium* detection (40%). A modified Poisson regression model is also considered more robust to outliers and avoids non-convergence issues compared to a log-binomial model for common binary outcomes (Chen *et al.* 2014). A backward variable selection process was used based on p -value and reduction in the Quasi-likelihood under the independence model criterion statistic (analogous to Akaike's Information Criterion). Variables with a p -value ≤ 0.05 were ultimately retained in the final model. Once the independent variables for the multivariable model were identified, a screening process for interaction terms among these variables was undertaken. Only interactions deemed plausible and relevant were assessed. Cook's distance, leverage, and residual plots were examined to identify influential points and assess model fit.

A classification tree, a type of machine learning method, was also created as an alternative approach to identify predictors of *Cryptosporidium* detection (Breiman *et al.* 1984). Compared to regression, these models have fewer restrictions on the type of data that can be included, better accommodate missing data, and are not confined to the implicit assumptions of regression. The splits may also suggest strategies for drinking water managers to tailor monitoring and interventions if they show adequate predictive accuracy. In building the tree, entropy was used to assess candidate splits for each node. As with the regression model, variables with >20% missing values were excluded; the remaining missing values were assigned to the most popular node. To help prevent overfitting, and to create a simpler tree, the tree was pruned back with cost-complexity pruning (Breiman *et al.* 1984) and a minimum leaf size of five. Cost-complexity analysis plots were used to select the final subtree. Ten-fold cross-validation was performed rather than hold-out validation due to the relatively small size of the dataset.

All analyses were performed using SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) with proc GENMOD procedure used for the Modified Poisson regression models and the HPSPLIT procedure used for the classification tree.

3. RESULTS AND DISCUSSION

3.1. Univariable analysis

Descriptive statistics, visual plots, and statistical test results for the 79 predictor variables are found in Tables S5–S9, with the findings summarized in Table 4.

In the ‘Well use and construction’ theme, three variables met the criteria (p -value ≤ 0.2) for retention in further analyses (Table 2 and Table S5). Shallower well depth, depth cased, and smaller saturated casing values were more likely among wells with a *Cryptosporidium* detection. Several of the well characteristic variables had a high percent of missing data ($>20\%$) and were excluded from further analysis; notably, grouted yes/no, grout material, percent casing grouted, percent grout saturated, annular space, and casing jointing method. Of these, a higher proportion of *Cryptosporidium*-positive wells was visually apparent among wells that were not grouted, constructed with B group (bentonite) grout material, and with larger annular space (Table S5).

In the ‘Aquifer characteristics, connectedness between aquifer and land surface’ theme, *Cryptosporidium*-positive wells were more likely to have porosity characteristics of unconsolidated sediments or fractured crystalline bedrock, higher land surface elevation, lower drawdown (which had a high percent of missingness), and water quality characteristics reflecting modern groundwater age (Table 2 and Table S6). These characteristics include detectable tritium, evidence for human impact based on elevated chloride/bromide ratios and evidence of rapid recharge based on fluctuating water quality parameters. While not statistically significant, the proportion of wells with detections increased from lowest, middle, to highest categories of geologic sensitivity (Table S6).

Several ‘Well capture zone and land use within capture zone’ variables met the criteria (p -value ≤ 0.2) for retention in further analyses (Table 2 and Table S7). *Cryptosporidium*-positive wells had lower levels of low, medium, and high development intensity within the 1-year travel time (1-yr TT) well capture zone. These development variables were not below the significance cut-point for 10-year travel time (10-yr TT), but the same trend was seen (Table S7). A higher proportion of *Cryptosporidium*-positive wells were located where ‘development is mostly agriculture’ (1-yr TT) versus not, and *Cryptosporidium*-positive wells had a higher percent of open water or wetland in the well capture zones, with both the 1- and 10-yr TT capture zones below the ≤ 0.2 p -value threshold. Wells with detections were also more likely to have larger runoff catchment areas and a higher percent of impervious surfaces in the runoff catchment area.

For the ‘Potential contaminant sources in the IWMZ’ theme, the infrequent presence of contamination sources within the IWMZ for several variables (right-censored at >200 ft) resulted in re-classification into ordinal categories for analysis (Table S8). While 61% of wells were missing sewage treatment information, a higher proportion of wells with detections were seen for gravity trenches compared to mound systems and pressurized septic systems (Table 2 and Table S8). Within the IWMZ, increasing distance to septic/sewage systems was associated with decreasing likelihood of detections. While not meeting the threshold for further analysis, wells with *Cryptosporidium* detections had lower mean and median design flows and were more common for septic versus municipal sewer types.

Several coefficients of variation (CVs) within the ‘Chemical and isotopic parameters’ theme had p -values ≤ 0.2 (Table 2 and Table S9). Higher bromide, chloride, chloride/bromide ratio, boron, nitrate, total organic carbon, specific conductance, and $\delta 2\text{H}$ CV values were seen among wells with *Cryptosporidium* detections.

3.2. Multivariable regression model

Variables meeting the $p < .20$ criteria in univariate analysis were included in regression analysis, with $p < .05$ used as a cut-point for retention in the final model. The final adjusted multivariable model is shown in Table 3. Increased runoff catchment area-percent impervious, nitrate CV, and bromide CV were significantly associated with increased risk of *Cryptosporidium* detection. Shallower well depth was also associated with higher risk of detection (i.e., increasing well depth showed a protective effect). Of note, saturated casing value (tertiles, with ‘missing’ as the fourth group) was significant in the model, but only when well depth was excluded, and vice-versa. Well depth was retained due to fewer missing data. Lower percent low-intensity development was also associated with higher risk of detection. In univariable analysis, the other development intensities (medium and high) showed the same direction of effect, suggesting *Cryptosporidium* risk may be related to generally undeveloped land.

3.3. Classification tree

Entering all predictor variables with p -values $< .20$ in univariable analysis into a classification tree model with a minimum leaf size of five, a tree with seven leaves best balanced cost complexity and misclassification. Depth of well, bromide CV, runoff

Table 2 | Risk factors for *Cryptosporidium* detection in univariable analysis with select visual plots ($n = 135$)^a

Theme	Increased likelihood of <i>Cryptosporidium</i> -positive wells with:
Well Use and Construction	Shallower well depth ($p=.032$) and depth cased ($p=.056$); and smaller saturated casing value ($p=.007$).
Aquifer Characteristics, Aquifer and Land Surface Connectedness	Higher elevation ($p=.120$); lower drawdown ($p=.056$) ^b ; modern groundwater age ($p=.117$); and primary unconsolidated or secondary porosity types ($p=.048$).
Well Capture Zone, Land Use in Capture Zone	Larger runoff catchment area ($p=.067$) and pct runoff catchment area impervious ($p=.077$); lower pct low intensity development (1-yr TT, $p=.054$), pct medium intensity development (1-yr TT $p=.179$), and pct high intensity development (1-yr TT: $p=.037$ and 10-yr TT: $p=.093$); higher pct of open water/wetland (1-yr TT: $p=.140$ and 10-yr TT: $p=.137$); and land use that is mostly agriculture (1 yr TT $p=.162$).
Potential Contaminant Sources in Inner Well Management Zone	Gravity trench sewage treatment compared to mound systems ($p=.012$) ^{b,d} ; closer distance to septic/sewage system ($p=.197$).
Chemical and Isotopic Parameters	Higher coefficients of variation for chloride ($p=.055$), bromide ($p=.067$), chloride/bromide ratio ($p=.096$), nitrate ($p=.021$), specific conductance ($p=.110$), boron ($p=.128$), total organic carbon ($p=.018$), and $\delta 2H$ ($p=.155$).

^aVariables with p -values $\leq .20$ were included in further analysis. Continuous data were compared between groups using the Mann-Whitney U or Kruskal-Wallis test. The Chi-square test was used to perform intergroup comparisons. The Cochran-Armitage trend test was used for ordinal variables. Due to space considerations, not all figures are shown. See Tables S5–S9 for full results.

^bA high percent of missing data prevent further testing of this variable in multivariable analysis.

^cAquifer porosity type: 1 = primary unconsolidated, 2 = primary consolidated within 50 feet of the land surface, and 3 = secondary.

^dOther sewage treatment types could not be tested due to small counts.

Table 3 | Adjusted relative risks (95% CIs) for *Cryptosporidium* detection from the modified Poisson regression model

Variable	Relative risk ^a	95% CI	p-Value
Well depth	0.998	0.997–1.00	.026
Runoff catchment area, percent impervious	1.01	1.00–1.02	.009
Percent low-intensity development (tertiles)	Tertile 1	1.73	1.01–2.96
	Tertile 2	1.33	0.79–2.26
	Tertile 3	REF	–
Nitrate CV	1.30	1.06–1.60	.012
Bromide CV	2.59	1.04–6.45	.041

^aN = 135. Relative risks are shown as exponentiated coefficients. Models adjusted for phase of study (Phases 1, 2, or both).

Table 4 | Classification tree output for *Cryptosporidium* detection**Variable importance**

Variable	Training		
	Relative importance	Importance	Count
Well depth	1.00	2.27	1
Bromide CV	0.99	2.26	1
Runoff catchment area, percent impervious	0.98	2.22	1
NO ₃ CV	0.96	2.18	1
Percent open water/wetland (TT)	0.71	1.61	1
Runoff catchment area	0.57	1.30	1

Fit statistics

	N leaves	Misclassification	Sensitivity	Specificity	AUC ^a
Model based	7	0.200	0.714	0.861	0.82
Cross-validation	7	0.440	0.446	0.633	

^aAUC = The area under the receiver operating characteristic curve. A model that fits the data perfectly would have an AUC of 1.

catchment area-percent impervious, and nitrate CV ranked comparably in importance, followed by percent open water/wetland (1-yr TT) and runoff catchment area (Table 4). The model incorrectly predicted the outcome for 20% of wells, with a higher misclassification rate in the cross-validation (44%), indicating that the model may not generalize well to new data. The model area under the curve (AUC) was 0.82, which is considered acceptable discrimination.

The first split was based on well depth, with wells <118 feet approximately twice as likely to have a *Cryptosporidium* detection (61% of wells) compared to 32% of wells ≥118 feet deep (Figure 2). For wells ≥118 feet, all wells with a runoff catchment area of ≥0.83 acres had a detection, but the sample size was small (n = 5). Of those with a runoff catchment area <0.83 acres, 86% of wells had detections when bromide CV was ≥32% (n = 7) compared to 23% of wells with a bromide CV <30%. Of these 79 wells, if nitrate CV was ≥130%, 83% had detections (n = 6) compared to those with nitrate CV <130%, in which 18% had detections.

Of wells <118 feet, all wells with a runoff catchment area ≥0.83 acres had detections, but the sample size is small (n = 5). Of wells with a runoff catchment area <0.83 acres and open water or wetland <1%, 27% had a detection compared to those with open water or wetland >1% in which 68% had detections.

3.4. Synthesis of findings

In a state-wide surveillance study, *Cryptosporidium* detection was unexpectedly common in Minnesota public supply wells. Adding to this concern, the incidence of cryptosporidiosis is reported to be increasing, with rates highest in northern Midwest states (CDC 2019). *Cryptosporidium*'s previously reported lack of correlation with coliform bacteria (Rose 1997; Stokdyk et al. 2020), low threshold for infectivity (DuPont et al. 1995), and high tolerance to chlorine (Betancourt & Rose 2004;

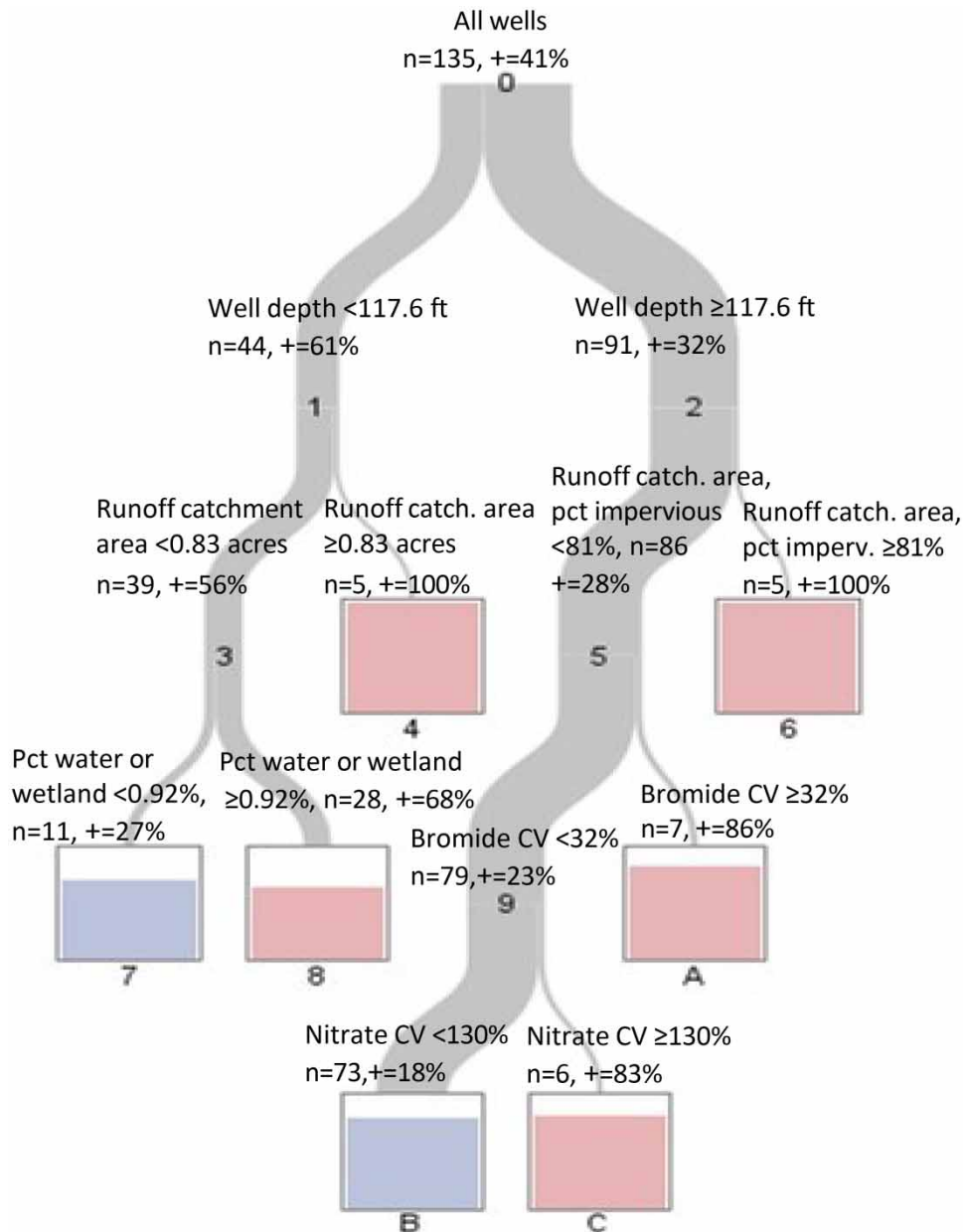


Figure 2 | Classification tree: risk factors for *Cryptosporidium* detection. Notes: += percent of wells with a *Cryptosporidium* detection. The color of the bar in each leaf node indicates the most frequent *Cryptosporidium* detection outcome (detection = red, no detection = blue) and represents the classification assigned to all observations in that node. The height of the bar indicates the proportion of wells in the node that have the most frequent outcome.

Fradette *et al.* 2022) make it critical to fill current gaps in knowledge regarding risk factors for *Cryptosporidium* in drinking water wells. Climate change also heightens the urgency of this work, as extreme precipitation events can intensify pollutant runoff and pathogen infiltration and subsequently increase the risk of waterborne disease outbreaks, including cryptosporidiosis (Curriero *et al.* 2001; Ikiroma & Pollock 2021). In addition, its large size may allow *Cryptosporidium* to serve as a bellwether for smaller pathogens such as bacteria and viruses. For these reasons, we explored potential risk factors for *Cryptosporidium* presence in public drinking supply wells.

Univariable analysis findings were considered independently in addition to serving as a prescreening for multivariable modeling. Certain physical and chemical factors across all five themes were found to significantly differ between wells with and without a *Cryptosporidium* detection in univariable analysis. This demonstrates that contamination risk cannot be solely

attributed to one type of risk factor, such as well construction characteristics. Shallower well and casing depths and lower saturated casing values were significant risk factors. Riskier aquifer characteristics included unconsolidated glacial sediments or bedrock aquifers dominated by secondary porosity and the presence of modern water based on tritium detection. An increase in the proportion of wells with detections with increasing geologic sensitivity was visually apparent as well. While [Stokdyk et al. \(2020\)](#) found that aquifer type and geologic sensitivity were not significant predictors of overall pathogen risk, when looking only at protozoa detections, sand and gravel aquifers and fractured crystalline rocks were noted as showing the greatest proportion of positive wells, with the same being said of the wells with the highest geologic sensitivity ratings. Indicators of human-impacted water quality based on elevated chloride/bromide ratios, and/or high temporal variability reflective of relatively rapid recharge, were also indicative of higher risk. In addition, smaller well drawdowns, suggestive of higher aquifer transmissivities, were identified as significant; however, hydraulic conductivity and aquifer thickness were also evaluated separately and not found to be significant factors. Similarly curious is the finding of higher land surface elevation being a factor in *Cryptosporidium* occurrence.

Our analyses also noted the importance of landform, land use, and land cover characteristics in relation to *Cryptosporidium* risk. For example, the absence of commercial, industrial, or residential development within the 1- or 10-year TT capture areas for wells, and/or the development within these areas consisting primarily of agricultural use, was related to elevated *Cryptosporidium* risk, as was the presence of open water or wetlands within these areas. Similarly, the presence of relatively large surface water catchments in the immediate vicinity of study wells, and the presence of a relatively large proportion of impermeable surfaces such as roadways and rooftops within these catchments, increased *Cryptosporidium* risk. These findings may reflect on the prevalence of potentially zoonotic *Cryptosporidium* species such as *C. parvum* in this study, rather than human-specific species, although the observations described above related to linkage with indicators of human-impacted water quality are at odds with this finding, as is the observation of higher risk associated with proximity and density of sewage treatment systems within the IWMZ, a finding similar to that described by [Borchardt et al. \(2021\)](#). Taken together, these sets of findings may reflect on the range of scenarios that may increase *Cryptosporidium* risk at wells, spanning the relative absence of development, where presumably zoonotic sources of *Cryptosporidium* predominate, to the presence of at least minimal development in terms of nearby septic systems or impervious surfaces that may impart associated indicators of human-impacted water quality such as elevated chloride/bromide ratios.

The final multivariable regression model included well depth, runoff catchment area, runoff catchment area-percent impervious, bromide CV, and nitrate CV as statistically significant predictors. Other general chemistry parameter CVs were significant in the multivariable model in the absence of each other, indicating redundancy, and the final model retained the CVs with the 'strongest' effects. Variability in these chemical parameters may reflect dynamic responses to recharge events ([Walsh et al. in press](#)) and/or other transient hydrologic factors such as well pumping. These insights can support monitoring efforts that include time-series or continuous data collection of well water chemistry to help better assess microbial risk. Each of these CV relationships was positive, with greater values indicating higher risk.

Classification trees are well suited to the study of environmental data because they can accommodate complex interactions among variables, allowing for the observation of moderating effects that might be missed when using traditional model-based approaches. Classification trees also provide easy to follow visualizations of the predictive process and the hierarchical importance of the variables from the top to the bottom of the tree, and therefore do not suffer from the 'black box' effect of more complicated machine learning approaches. The ability of classification trees to efficiently segment a 'population' of wells into meaningful subsets can help drinking water managers identify potentially high-risk wells, or their inverse, so that resources (e.g., monitoring) can be targeted accordingly. For this reason, we intentionally created a simple tree model using cost complexity pruning and limiting the number of nodes to make the tree useful for public health decision making. The cross-validation fit showed higher specificity than sensitivity, with wells less likely to have detections having 1) well depth >118 feet and a runoff catchment area-percent impervious <81%, unless bromide CV was >32%. For shallower wells (<118 feet), smaller runoff catchment areas with lower percent open water or wetland decreased the likelihood of detections. While classification trees and regression models are different algorithms, we found that both the regression and classification tree models identified nitrate and bromide CVs, well depth, and runoff catchment area (or runoff catchment area-percent impervious) as influential in *Cryptosporidium* occurrence, increasing confidence in these findings.

The protective effects of increasing well depth, less potential for surface water runoff around the well, and the absence of high bromide variability reflective of rapid or changing recharge conditions make intuitive hydrogeologic sense and are generally substantiated in groundwater literature. For example, numerous studies have described correlation between shallower

well depths and microbial contamination (e.g., [Gonzales 2008](#); [Maran et al. 2016](#); [O'Dwyer et al. 2018](#)). While none of these settled on the same primary cutoff depth noted here (118 feet), those studies were focused on microbial contaminants other than *Cryptosporidium* and likely included wells of differing characteristics, such as depth ranges and construction features, than those in this study. Variability in chemical parameters has been recognized as reflective of rapid recharge and degraded water quality, but again, not specific to *Cryptosporidium* risk or to bromide or nitrate variability. For example, [Jacobson & Langmuir \(1974\)](#) found CVs of many ionic parameters >15% to be reflective of rapid groundwater flow conditions in conduit karst, with slightly lower values seen in diffuse flow systems, and [Dhar et al. \(2008\)](#) found that wells deeper than 30 m showed significantly less temporal chemical variability than shallower wells. [Bexfield & Jurgens \(2014\)](#) found that seasonal variability in water quality at public supply wells was influenced by well usage patterns, and [Aisopou et al. \(2015\)](#) found that pesticide concentrations in production wells were dependent on variables such as pumping rate and the hydrogeology of the aquifer. [Stokdyk et al. \(2019\)](#) also looked at variability in chemical and isotopic parameters for the same data set used in this analysis on the basis that such variability likely reflected rapid recharge and microbial risk. However, the authors found that *Cryptosporidium*-positive wells were no more likely to show evidence of rapid recharge or surface water influence, based on the evaporative signature of water isotopes, than wells that lacked detections. These conclusions are likely at odds with the current analysis because the 2019 analysis utilized arbitrarily determined threshold values for establishing variability, while including the evaporative signature analysis, whereas the current analysis simply looked at the full range of observed values for all chemical and isotopic parameters to look for association with *Cryptosporidium* occurrence. Relationships between surface water runoff catchment area and impervious surfaces and pathogen risk are less substantiated in the literature beyond the effects of well flooding ([Andrade et al. 2018](#)), and the observation that extreme climatic events have been associated with waterborne disease outbreaks more generally ([Curriero et al. 2001](#); [Auld et al. 2004](#); [Cann et al. 2013](#)), implying possible correlation with runoff mechanisms. While apparently novel, the observations described in this study about association with surface water runoff make intuitive sense, given that the flow or ponding of surface water that might be contaminated with *Cryptosporidium* in the vicinity of drinking water wells could exploit macropores or other fast flow pathways into the subsurface.

While this study focused on public wells, there are important takeaways for private wells, which are estimated to serve 17% of the U.S. population ([Murray et al. 2021](#)). The observed importance of well depth as a predictor of *Cryptosporidium* risk is concerning given that private wells are generally shallower than public wells. For example, the average private well depth in Minnesota is 138 feet ($n = 415,638$) compared to 179 feet ($n = 20,171$) for public wells (Minnesota Well Index database, as of 1/12/2023). In addition, full-length grouting is not required for private wells, and well siting considerations may be less rigorous than those used for public wells, or not considered at all for wells drilled before the Minnesota Well Code took effect in 1974. Finally, private wells have no water quality testing requirements after the time of construction in Minnesota as well as most other states and are unlikely to be sampled repeatedly or for the parameters deemed useful in this analysis, outside of special studies.

There are several unique features of this study. First, *Cryptosporidium* occurrence in wells is an understudied area of research, making this work an important contribution. Secondly, a relatively large number of wells were monitored multiple times (typically 6 times, up to 12) and this greater sampling frequency was important since positive wells were often found to have a few positive detections accompanied by multiple negative results. The application of multiple data analysis approaches involving a wide range of potential risk factors is also a strength.

There are also some limitations to this analysis. First, the focus of this work was on *Cryptosporidium* occurrence, not concentration, which may have a unique set of risk factors. Second, the analysis did not consider potential meteorological predictors such as preceding precipitation events, which have been shown to be influential. In a separate Minnesota study, springtime, the season with the highest amount of rainfall in Minnesota, was associated with the greatest frequency of microbial contamination ([MDH 2023](#), in press). However, the longitudinal collection strategy was designed to capture results across seasons. Certain variables could not be adequately assessed due to a high frequency of missing data. Univariable analyses suggest some of these variables, such as percent grout saturated, annular space, drawdown, and septic system design flow, may be important predictors of *Cryptosporidium* detection. However, the lack of available data indicates that these characteristics are not often available to drinking water programs, making their utility in identifying at-risk wells limited from a practical standpoint.

There are reasons why these findings should be interpreted with caution. Both multivariable regression models and tree models can be unstable and result in overfitting when the number of observations is small relative to the number of predictor

parameters. We also do not know whether these findings are generalizable to other areas within or outside the U.S.; additional studies conducted in a variety of settings are needed. Despite these caveats, the results suggest that more groundwater systems should be monitored for *Cryptosporidium* based on the risk and protection factors identified here and highlighted in other relevant reports. Based on the findings of this work, MDH intends to bolster the predictive outcome of its own existing well vulnerability scoring rubrics by adding additional variables (such as bromide variability and runoff/impervious area around wells), assigning heavier weight to variables that passed more rigorous tests, and using specific cutoff values for the parameters identified here.

4. CONCLUSIONS

Risk of *Cryptosporidium* occurrence in wells was associated with multiple anthropogenic and natural factors, suggesting both a wide array of influential factors and diverse range of mammalian hosts of the different species of this organism. Shallower well depth and depth cased, lower saturated casing values, potential for surface water runoff around the well, and variability in chemical parameters reflective of rapid or changing recharge, along with certain land use and land characteristics, were consistent predictors of occurrence across data analysis methods. Drinking water programs can work to minimize the risk of cryptosporidiosis by considering the factors identified here when developing predictive scoring rubrics, well monitoring strategies, and well construction and siting plans. The findings should also be considered when assessing the adequacy of current drinking water regulatory standards for *Cryptosporidium*.

ACKNOWLEDGEMENTS

Funding was provided by the Minnesota Clean Water, Land, and Legacy Amendment Fund. The authors wish to thank the public water suppliers who participated in the study and the MDH team responsible for field duties including Dane Huber, Jared Schmaedeke, Trisha Sisto, Mike Sutliff, and Nathan Gieske. Alycia Overbo of MDH provided communications support. Laboratory assistance and insight were provided by Aaron Firnstahl (USGS), Joel Stokdyk (USGS), Susan Spencer (USDA-ARS retired) and Mark Borchardt (USDA-ARS retired).

AUTHOR CONTRIBUTIONS

James Walsh contributed to the conceptualization, project administration, investigation, writing the original draft, and funding acquisition. Deanna Scher contributed to the formal analysis, visualization, and writing the original draft. Jane de Lambert contributed to the conceptualization, data curation, writing the original draft, and also reviewed and edited the manuscript. Anita Anderson contributed to the conceptualization, writing the original draft, and also reviewed and edited the manuscript.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Aisopou, A., Binning, P. J., Albrechtsen, H. J. & Bjerg, P. L. 2015 Modeling the factors impacting pesticide concentrations in groundwater wells. *Groundwater* **53** (5), 722–736. doi:10.1111/gwat.12264.
- Andrade, L., O'Dwyer, J., O'Neill, E. & Hynds, P. 2018 Surface water flooding, groundwater contamination, and enteric disease in developed countries: a scoping review of connections and consequences. *Environ. Pollut.* **236**, 540–549. doi:10.1016/j.envpol.2018.01.104.
- Auld, H., MacIver, D. & Klaassen, J. 2004 Heavy rainfall and waterborne disease outbreaks: The Walkerton example. *J. Toxicol. Environ. Health* **67** (20–22), 1879–1887. doi:10.1080/15287390490493475.
- Betancourt, W. Q. & Rose, J. B. 2004 Drinking water treatment processes for removal of *Cryptosporidium* and *Giardia*. *Vet. Parasitol.* **126** (1–2), 219–234. doi:10.1016/j.vetpar.2004.09.002.
- Bexfield, L. M. & Jurgens, B. C. 2014 Effects of seasonal operation on the quality of water produced by public-supply wells. *Groundwater* **52** (Suppl 1), 10–24. doi:10.1111/gwat.12174.

- Borchardt, M. A., Stokdyk, J. P., Kieke Jr, B. A., Muldoon, M. A., Spencer, S. K., Farnstahl, A. D., Bonness, D. E., Hunt, R. J. & Burch, T. R. 2021 Sources and risk factors for nitrate and microbial contamination of private household wells in the fractured Dolomite aquifer of Northeastern Wisconsin. *Environ. Health Perspect.* **129** (6), 67004. doi:10.1289/EHP7813.
- Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. 1984 *Classification and Regression Trees*, (1st edn.). Routledge, New York. doi:10.1201/9781315139470.
- Burch, T. R., Stokdyk, J. P., Rice, N., Anderson, A. C., Walsh, J. F., Spencer, S. K., Farnstahl, A. D. & Borchardt, M. A. 2022 Statewide quantitative microbial risk assessment for waterborne viruses, bacteria, and protozoa in public water supply wells in Minnesota. *Environ. Sci. Technol.* **56** (10), 6315–6324. doi:10.1021/acs.est.1c06472.
- Cann, K. F., Thomas, D. R., Salmon, R. L., Wyn-Jones, A. P. & Kay, D. 2013 Extreme water-related weather events and waterborne disease. *Epidemiol. Infect.* **141** (4), 671–686. doi:10.1017/S0950268812001653.
- Centers for Disease Control and Prevention. 2019 *Cryptosporidiosis NNDSS Summary Report for 2019*. Available from: <https://www.cdc.gov/healthywater/surveillance/cryptosporidium/cryptosporidium-2019.html> (accessed 7 August 2023).
- Centers for Disease Control and Prevention. 2023 *Parasites – Cryptosporidium (also known as ‘Crypto’): General Information for the Public*. Available from: <https://www.cdc.gov/parasites/cryptosporidium/general-info.html> (accessed 26 June 2023).
- Chen, W., Shi, J., Qian, L. & Azen, S. P. 2014 Comparison of robustness to outliers between robust Poisson models and log-binomial models when estimating relative risks for common binary outcomes: a simulation study. *BMC Med. Res. Methodol.* **14**, 82. doi:10.1186/1471-2288-14-82.
- Chique, C., Hynds, P. D., Andrade, L., Burke, L., Morris, D., Ryan, M. P. & O’Dwyer, J. 2020 *Cryptosporidium* spp. in groundwater supplies intended for human consumption – A descriptive review of global prevalence, risk factors and knowledge gaps. *Water Res.* **176**, 115726. doi:10.1016/j.watres.2020.115726.
- Curriero, F. C., Patz, J. A., Rose, J. B. & Lele, S. 2001 The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. *Am. J. Public Health* **91** (8), 1194–1199. doi:10.2105/ajph.91.8.1194.
- Dhar, R. K., Zheng, Y., Stute, M., van Geen, A., Cheng, Z., Shanewaz, M., Shamsudduha, M., Hoque, M. A., Rahman, M. W. & Ahmed, K. M. 2008 Temporal variability of groundwater chemistry in shallow and deep aquifers of Araihaazar, Bangladesh. *J. Contam. Hydrol.* **99** (1–4), 97–111. doi:10.1016/j.jconhyd.2008.03.007.
- DuPont, H. L., Chappell, C. L., Sterling, C. R., Okhuysen, P. C., Rose, J. B. & Jakubowski, W. 1995 The infectivity of *Cryptosporidium parvum* in healthy volunteers. *N. Engl. J. Med.* **332** (13), 855–859. doi:10.1056/NEJM199503033321304.
- Fradette, M. S., Culley, A. I. & Charette, S. J. 2022 Detection of *Cryptosporidium* spp. and *Giardia* spp. in environmental water samples: a journey into the past and new perspectives. *Microorganisms* **10** (6), 1175. doi:10.3390/microorganisms10061175.
- Gonzales, T. R. 2008 The effects that well depth and wellhead protection have on bacterial contamination of private water wells in the Estes Park Valley, Colorado. *J. Environ. Health* **71** (5), 17–23.
- Hancock, C. M., Rose, J. B. & Callahan, M. 1998 *Crypto and Giardia in U.S. groundwater*. *J. AWWA* **90** (3), 58–61. doi:10.1002/j.1551-8833.1998.tb08396.x.
- Hynds, P. D., Thomas, M. K. & Pintar, K. D. 2014 Contamination of groundwater systems in the U.S. and Canada by enteric pathogens, 1990–2013: A review and pooled-analysis. *PLoS One* **9** (5), e93301. doi:10.1371/journal.pone.0093301.
- Ikiroma, I. A. & Pollock, K. G. 2021 Influence of weather and climate on cryptosporidiosis – A review. *Zoonoses Public Health* **68** (4), 285–298. doi:10.1111/zph.12785.
- Jacobson, R. L. & Langmuir, D. 1974 Controls on the quality variations of some carbonate spring waters. *J. Hydrol.* **23** (3–4), 247–265. doi:10.1016/0022-1694(74)90006-7.
- Landon, M. K., Delin, G. N., Komor, S. C. & Regan, C. R. 2000 Relation of pathways and transit times of recharge water to nitrate concentrations using stable isotopes. *Groundwater* **38** (3), 381–395.
- Landwehr, J. M. & Coplen, T. B. 2004 Line-conditioned excess: A new method for characterizing stable hydrogen and oxygen isotope ratios in hydrologic systems. In: *Proceedings of Isotopes in Environmental Studies*, Monaco, 25–29 October 2004, Chapter IAEA-CN-118/56 (IAEA, ed.). Available from: https://www-pub.iaea.org/MTCD/Publications/PDF/CSP_26_web.pdf
- Maran, N. H., Crispim, B. D., Jahnn, S. R., Araújo, R. P., Grisolia, A. B. & Oliveira, K. M. 2016 Depth and well type related to groundwater microbiological contamination. *Int. J. Environ. Res. Public Health* **13** (10), 1036. doi:10.3390/ijerph13101036.
- McNutt, L. A., Wu, C., Xue, X. & Hafner, J. P. 2003 Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am. J. Epidemiol.* **157** (10), 940–943. doi:10.1093/aje/kwg074.
- MDH. 2023 *MDH Pathogen Project Recharge Monitoring Study Final Report*. Minnesota Department of Health, St Paul, MN. Available from: <https://www.health.state.mn.us/communities/environment/water/docs/swp/pathogen.pdf>.
- Murphy, H. M., Prioleau, M. D., Borchardt, M. A. & Hynds, P. D. 2017 Epidemiological evidence of groundwater contribution to global enteric disease, 1948–2015. *Hydrogeol. J.* **25** (4), 981–1001. doi:10.1007/s10040-017-1543-y.
- Murray, A., Hall, A., Weaver, J. & Kremer, F. 2021 Methods for estimating locations of housing units served by private domestic wells in the United States applied to 2010. *J. Am. Water Resour. Assoc.* **57** (5), 1–16. doi:10.1111/1752-1688.12937.
- O’Dwyer, J., Hynds, P. D., Byrne, K. A., Ryan, M. P. & Adley, C. C. 2018 Development of a hierarchical model for predicting microbiological contamination of private groundwater supplies in a geologically heterogeneous region. *Environ. Pollut.* **237**, 329–338. doi:10.1016/j.envpol.2018.02.052.

- Rose, J. B. 1997 Environmental ecology of *Cryptosporidium* and public health implications. *Ann. Rev. Public Health* **18**, 135–161. doi:10.1146/annurev.publhealth.18.1.135.
- Smith C. M. & Hill, V. R. 2009 Dead-end hollow-fiber ultrafiltration for recovery of diverse microbes from water. *Appl. Environ. Microbiol.* **75**, 5284–5289.
- Stokdyk, J. P., Firnstahl, A. D., Spencer, S. K., Burch, T. R. & Borchardt, M. A. 2016 Determining the 95% limit of detection for waterborne pathogen analyses from primary concentration to qPCR. *Water Res.* **96**, 105–113.
- Stokdyk, J. P., Spencer, S. K., Walsh, J. F., de Lambert, J. R., Firnstahl, A. D., Anderson, A. C., Rezania, L. W. & Borchardt, M. A. 2019 *Cryptosporidium* incidence and surface water influence of groundwater supplying public water systems in Minnesota, USA. *Environ. Sci. Technol.* **53** (7), 3391–3398. doi:10.1021/acs.est.8b05446.
- Stokdyk, J. P., Firnstahl, A. D., Walsh, J. F., Spencer, S. K., de Lambert, J. R., Anderson, A. C., Rezania Jr, L. W., Kieke, B. A. & Borchardt, M. A. 2020 Viral, bacterial, and protozoan pathogens and fecal markers in wells supplying groundwater to public water systems in Minnesota, USA. *Water Res.* **178** (1), 115814. doi:10.1016/j.watres.2020.115814.
- Watier-Grillot, S., Costa, D., Petit, C., Razakandrainibe, R., Larréché, S., Tong, C., Demont, G., Biletorte, D., Mouly, D., Fontan, D., Velut, G., Le Corre, A., Beauvir, J. C., Mérens, A., Favennec, L. & Pommier de Santi, V. 2022 *Cryptosporidiosis* outbreaks linked to the public water supply in a military camp, France. *PLoS Negl. Trop. Dis.* **16** (9), e0010776. doi:0.1371/journal.pntd.0010776.
- Zou, G. 2004 A modified Poisson regression approach to prospective studies with binary data. *Am. J. Epidemiol.* **159** (7), 702–706.

First received 15 November 2023; accepted in revised form 23 January 2024. Available online 5 February 2024