

Total coliform and *Escherichia coli* contamination in rural well water: analysis for passive surveillance

Jesse Invik, Herman W. Barkema, Alessandro Massolo, Norman F. Neumann and Sylvia Checkley

ABSTRACT

With increasing stress on our water resources and recent waterborne disease outbreaks, understanding the epidemiology of waterborne pathogens is crucial to build surveillance systems. The purpose of this study was to explore techniques for describing microbial water quality in rural drinking water wells, based on spatiotemporal analysis, time series analysis and relative risk mapping. Tests results for *Escherichia coli* and coliforms from private and small public well water samples, collected between 2004 and 2012 in Alberta, Canada, were used for the analysis. Overall, 14.6 and 1.5% of the wells were total coliform and *E. coli*-positive, respectively. Private well samples were more often total coliform or *E. coli*-positive compared with untreated public well samples. Using relative risk mapping we were able to identify areas of higher risk for bacterial contamination of groundwater in the province not previously identified. Incorporation of time series analysis demonstrated peak contamination occurring for *E. coli* in July and a later peak for total coliforms in September, suggesting a temporal dissociation between these indicators in terms of groundwater quality, and highlighting the potential need to increase monitoring during certain periods of the year.

Key words | Canada, coliforms, *Escherichia coli*, geographic information systems, groundwater, relative risk mapping

Jesse Invik
Alessandro Massolo
Sylvia Checkley (corresponding author)
 Department of Ecosystem and Public Health,
 Faculty of Veterinary Medicine,
 University of Calgary,
 2500 University Dr., Calgary,
 AB, Canada T2N 4N1
 E-mail: slcheckl@ucalgary.ca

Jesse Invik
Sylvia Checkley
 Alberta Provincial Laboratory for Public Health,
 Calgary Laboratory Site,
 3030 Hospital Dr. NW, Calgary,
 AB, Canada T2N 4W4

Herman W. Barkema
 Department of Production Animal Health, Faculty
 of Veterinary Medicine,
 University of Calgary, Calgary,
 AB, Canada T2N 4N1

Norman F. Neumann
 School of Public Health,
 University of Alberta,
 3-300 Edmonton Clinic Health Authority, 11405-87
 Ave, Edmonton,
 AB, Canada T6G 1C9
 and
 Alberta Provincial Laboratory for Public Health,
 Edmonton Laboratory Site,
 8440-112 St, Edmonton,
 AB, Canada T6G 2J2

INTRODUCTION

Surveillance of water for microbiological pathogens has traditionally involved the use of indicator organisms (Standridge 2008; World Health Organization 2011). Total coliforms and *Escherichia coli* have been used as indicators of water quality worldwide (Gleeson & Gray 1996; World Health Organization 2011). The World Health Organization recommends *E. coli* as an 'essential parameter' of minimum water monitoring (World Health Organization 2011). Protection of drinking water requires a multi-barrier approach, including monitoring and management, legislation and guidelines, empowering and

informing the public and research for new technological solutions (Federal-Provincial-Territorial Committee on Drinking Water and CCME Water Quality Task Group 2004). Not everyone is subject to drinking water legislation; people living in rural areas often depend on groundwater, most often untreated, for their drinking water (Summers 2010), potentially putting them at greater risk for waterborne illness than their urban counterparts (Galanis *et al.* 2014).

In Canada, regulations regarding drinking water are overseen by the provincial governments, and thus vary by

province, especially for small public and private systems. For instance, in British Columbia small systems with two or more connections fall under regulations, but in Quebec systems that serve 20 individuals or less are not regulated (Cook *et al.* 2013). In the United States, the Environmental Protection Agency (EPA) regulates public drinking water systems, which are defined as those systems serving 15 connections or 25 individuals (United States Environmental Protection Agency 2015). Private drinking water systems (accessed by approximately 15% of the US population) are not regulated by the EPA (United States Environmental Protection Agency 2012). Consequently, many people in North America and around the world consume groundwater from private wells for which public health is not protected through legislation. Those consuming groundwater without regular testing may be at risk for waterborne disease.

The testing provided by provincial or state laboratories can be used as a foundation of a surveillance plan for microbial water quality, but baseline levels including seasonality and trends need to be established for comparison to future levels. In addition, it is important to understand the current spatial distributions of contamination in order to effectively interpret potential outbreak data (Hay *et al.* 2013) and determine how future climatic changes may alter the distribution of waterborne pathogen risks and outbreaks (Bezirtzoglou *et al.* 2011; Galway *et al.* 2015). Passive data collection, often referred to as passive surveillance, is an economically advantageous method of sampling a large population, or developing a large dataset over a number of years, where active collection may not be feasible or affordable. Although passive data collection can have its drawbacks, such as self-selection bias or incomplete sampling, it is reported to have excellent sensitivity when the dataset is large enough, even if disease prevalence is low (Craighead *et al.* 2015).

With the advance of more user-friendly geographical information systems in the late 1990s, spatial analysis of epidemiological data has become a key tool for visualizing disease processes spatially, tracing the sources of disease and identifying areas with greater risk of disease (Stevenson *et al.* 2008). Spatial analysis methods in epidemiology include simple spatial visualization of health indicator patterns, local and global disease/pathogen cluster detection

methods, spatial interpolation, spatial risk assessment and regression models which incorporate spatial dependency (Stevenson *et al.* 2008). These methods have been applied to water contamination research worldwide. The city of Puri, India, used point sampling of water wells and interpolation to create contour maps of groundwater levels in pre- and post-monsoon conditions and identify the seasonal patterns and distribution of bacterial and chemical contaminants (Vijay *et al.* 2011). The results allowed the authors to make several suggestions to reduce future water contamination. In Canada, a 2013 study of Ontario private well water used a spatial scan statistic methodology employing a circular window to identify spatial clusters of *E. coli*-positive wells (Krolik *et al.* 2013). Greater Vancouver, British Columbia used a number of variables including intrinsic aquifer susceptibility, well location records, digital elevation models, land use data and known groundwater contamination sites to create a risk map for water sources in the area. This project also produced a relative risk map, but this map was based on potential risk factors, not on actual contamination outcomes, and focused on a much smaller geographical area (Simpson *et al.* 2014).

Relative risk maps, also referred to as excess rate maps, are used to demonstrate areas of higher or lower risk for disease (Anselin *et al.* 2010). Using the overall mean rate of disease for a large region, an expected rate for smaller regions within the large region, such as counties, can be calculated based on the population in each county. The ratio of expected versus actual cases allows a measure of relative risk in each county compared to neighbouring counties (Anselin *et al.* 2010). This methodology has been used to identify areas at risk for gastrointestinal illness in Northern Canada (Pardhan-Ali *et al.* 2012), and *Cryptosporidium* spp. contamination of surface water in Ireland (Samadder *et al.* 2010). Empirical Bayesian smoothing allows for correction of raw numbers in geographical areas with small populations, which can lead to misleading rates (Owusu-Edusei & Owens 2009).

Time series analyses are techniques often used in epidemiology, as well as a number of other disciplines, not only to track trends over time, but also to model future outcomes based on current and past occurrences (Shumway & Stoffer 2006). Time series analysis was recently used to model the

impact of hydroclimatic variables on waterborne gastrointestinal illness in British Columbia, Canada (Galway et al. 2015).

The objectives of this study were to investigate the use of relative risk mapping and time series analysis to establish baseline levels of contamination of rural groundwater with *E. coli* and total coliforms in the province of Alberta, Canada as a case study, and to explore the use of passive collection of voluntary water sample submissions as a tool for continued water surveillance activities. Specifically, we aimed to: (1) use spatiotemporal techniques to detect patterns in passively collected water contamination data; (2) test if patterns of contamination were spatially and temporally structured, and to what extent; (3) describe methodology for determining baseline levels of contamination and seasonality, as well as areas of greater or lower risk using spatiotemporal analysis and relative risk mapping techniques.

METHODS

Data sources

The study area included the entire province of Alberta, which is over 660,000 square kilometres and is located in Western Canada. The southern border of Alberta follows the 49th parallel and the northern border follows the 60th parallel. The eastern border with the province of Saskatchewan is delineated by the 110th meridian west, and the western border with the province of British Columbia is delineated by the 120th meridian from the north down to the continental divide, and then the border trends eastward following the divide. The province has a population of over four million people, representing 12% of the population of Canada (Statistics Canada 2014). Water submission data included 179,623 test results for *E. coli* and total coliforms for the years 2004–2012 for Alberta, Canada. Submissions were from rural well water samples (both small public systems and private wells). Testing was performed by the Alberta Provincial Laboratory for Public Health (ProvLab) (Calgary, AB, Canada) and accessed using the Data Integration for Alberta Laboratories (DIAL) tool, a web-based surveillance tool developed by ProvLab. ProvLab is an

ISO 17025 accredited laboratory for analysis of microbiological water.

Re-samples, samples collected for quantitative analysis after a positive initial test, were not included in this study. All water samples were collected by the well owners, and inclusion of name, address and location information accompanying the samples was left to the discretion of the person providing the samples. This information was handwritten by the sample submitter on standard provincial water requisition forms at the time of submission and subsequently entered into a computer system by the receiving technologist. If the sample was positive, the local public health agency was contacted by ProvLab and it was their responsibility to inform the well owner/overseer and provide information concerning decontamination and further testing as per Alberta Health and Wellness's *Environmental Public Health Field Manual* (Technical Advisory Committee on Safe Drinking Water 2004).

Water testing

Water was tested using a presence/absence enzyme substrate test for *E. coli*/total coliforms (Colilert® IDEXX, Westbrook, ME, USA) according to the manufacturer's protocol (IDEXX Laboratories 2013). One hundred mL specimens of water were incubated with the Colilert® product for 24 hours at 35 ± 0.5 °C. Water samples collected >24 hours before delivery to the laboratory were not analysed ($n = 429$, 0.2%). A further 67 (0.04%) samples lacked results due to submission or technical errors (too small volume, poor specimen quality or laboratory error). After incubation, the water sample was examined under natural light for a colour change from clear to yellow caused by metabolization of ortho-nitrophenyl- β -galactoside (ONPG) by β -galactosidase, indicating a total coliform positive test. If a colour change occurred, the sample was examined under ultraviolet light for fluorescence caused by metabolization of 4-methylumbelliferyl-beta-D-glucuronide (MUG) by β -glucuronidase, indicating an *E. coli*-positive test with a sensitivity of 1 colony forming unit per 100 mL (IDEXX Laboratories 2013). Facilities that had positive tests were encouraged to re-submit samples for quantitative analysis (re-submissions following positive tests were not included

in this study). Homeowners or private facility water operators are encouraged to submit multiple samples over the course of a year, aligning with the recommendations from Health Canada (Health Canada 2013).

Geolocation

Geographical coordinates of the submission data were derived from the Alberta Township Survey (ATS) System, a system for locating parcels of land in Alberta (Alberta Environment and Parks 2010). Each parcel is located by the closest meridian on its eastern side (the 4th, 5th or 6th), as well as its range, township, section and quarter section. This information allows a parcel of land to be georeferenced to a resolution of 1 quarter section ($\sim 800 \times 800$ m or 0.65 km²) (Alberta Environment and Parks 2010). In addition, Alberta Health Services (the government administrative body for health in the Province of Alberta) has divided the province into nine geographical health regions and this information was also used for mapping purposes.

Frequency of contamination, overall and by water source

Samples were categorized as being submitted by a private landowner or by a public unregulated system. Public systems included in this study were defined according to Alberta Environment and Parks as those that are not regulated by this ministry, and included non-transient systems with <15 connections and transient systems such as campgrounds and community halls (Alberta Environment 2009). In addition, samples were classified as having complete or non-existent/incomplete/invalid ATS geolocation data. Differences in contamination occurrence with *E. coli* and total coliforms between public and private wells while controlling for geolocation were examined using Cochran–Mantel–Haenszel (CMH) analysis (Fidalgo & Madeira 2008) using WinEpiscope 2.0 (Thrusfield et al. 2001). Frequency of occurrence of *E. coli* and/or total coliforms is reported rather than prevalence, as the denominator does not represent all wells at risk in the province, but rather all wells tested by ProvLab from voluntary submissions.

Repeat testing

Among those samples geolocated by ATS to the quarter section, the occurrence of repeated tests (multiple tests in the same quarter section in a single year) was examined, overall and according to water source, public and private. The distribution of water tests per year for public and private wells was examined using the Mann–Whitney U test using R (version 2.14.0, R Development Core Team 2011).

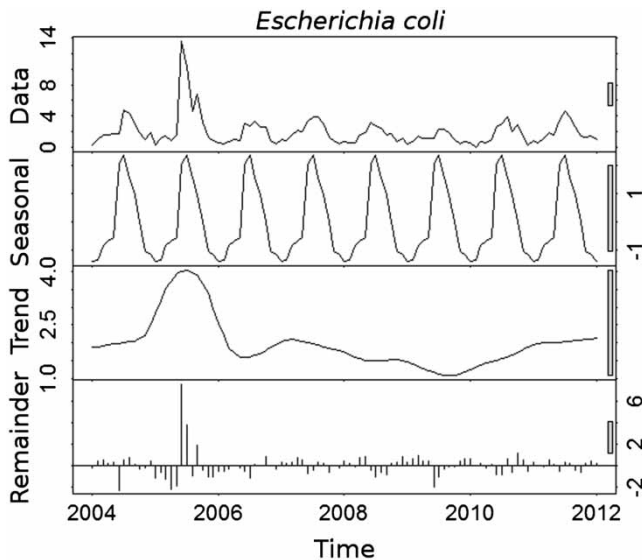
In order to mitigate the bias presented by the repeated testing of these data for the remainder of the analysis, the tests were aggregated by quarter section. One positive test within the quarter section during a month was counted as a positive outcome for that quarter section month.

Spatiotemporal analysis

The distribution of submissions and positive tests over day of the week, month and year was examined using time series and seasonal trend loess (STL) decomposition (Hafen et al. 2009). With the exception of days of the week, all time series analyses were based on data aggregated to quarter section and month. Differences between submission rates per day of the week were examined with Pearson's chi-square test using R (version 2.14.0, R Development Core Team 2011). To examine seasonal variation, a time series was created for days of the week, months and years over the study period (2004–2012). We created these time series by ordering the *E. coli*-positive, total coliform-positive and total number of tests performed by equally spaced time intervals in Microsoft Excel and graphing the results. For the days of the week and months time series, data for all years was aggregated into the appropriate day of the week or month of the year. The data were also divided into three regions (north, central and south) based on Alberta Health Services administrative areas, and a separate time series was created for each region. These regions were also chosen to represent the latitude and climate gradient running north to south. Edward's test for seasonality was performed using WINPEPI version 11.18 (Abramson 2011) and peak dates were determined for the entire province and for the three regions (Table 1). STL decomposition (Figures 1 and 2) was performed using R (version 2.14.0, R Core Development Team 2011) to decompose the time

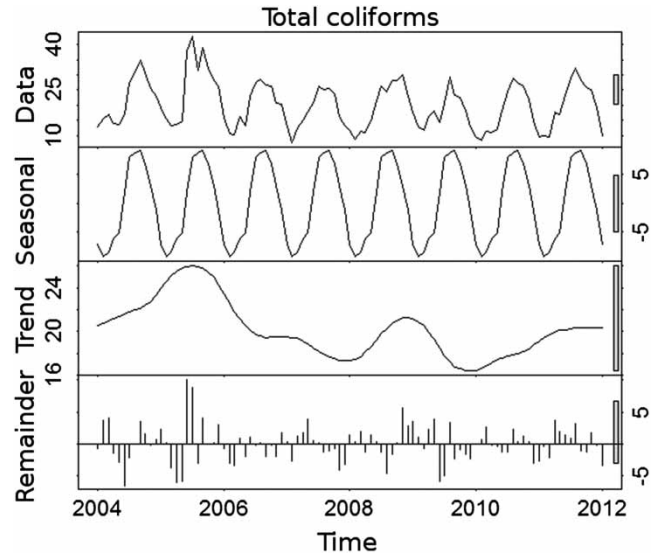
Table 1 | Edward's test of seasonality on water testing for total coliforms and *Escherichia coli* for north, central and southern regions of Alberta from 2004 to 2012

	χ^2	df	p-value	Peak date
<i>Escherichia coli</i>				
South	473.45	2	<0.001	July 23
Central	261.25	2	<0.001	July 21
North	94.87	2	<0.001	August 6
Total coliforms				
South	641.34	2	<0.001	August 18
Central	1,005.28	2	<0.001	September 6
North	358.79	2	<0.001	September 12

**Figure 1** | STL decomposition for *Escherichia coli* demonstrating raw data (first row), seasonal pattern (second row), trend over all years (third row) and remainder, the residuals from the fit of seasonal plus trend (fourth row) for private and public well water, 2004–2012, Alberta, Canada.

series into components: seasonal, trend and residuals, using a locally weighted non-parametric regression (Cleveland *et al.* 1990).

Two different sets of maps of water contamination patterns in Alberta were created using ArcGIS (version 10.1, ESRI 2012). The first set was prepared with sample level data aggregated to health region (Alberta's health regions have since been amalgamated), with percentage of *E. coli* and total coliform-positive wells (Figure 3). The second set of maps, relative risk maps, was prepared using the ATS geolocated data, aggregated by quarter section and month (Figure 4). The ATS geolocated data were treated as point

**Figure 2** | STL decomposition for total coliforms demonstrating raw data (first row), seasonal pattern (second row), trend over all years (third row) and remainder, the residuals from the fit of seasonal plus trend (fourth row) for private and public well water, 2004–2012, Alberta, Canada.

data using the centroids of the quarter sections. The point data were then aggregated to polygons small enough to preserve as much of the geographic location as possible while still providing confidentiality for regions with only a few wells. Polygons were based loosely on the ATS grid system and the average size was 13,266 km², with a standard deviation of 2,972 km².

Using aggregated data areas with small numbers of observations can produce misleading results (Cressie 1995). We used empirical Bayesian smoothing to derive confidence from areas with larger populations and adjust observations in areas with smaller populations towards the global mean. Empirical Bayesian smoothing of the crude proportions of *E. coli* and total coliform contamination in the polygons was performed using GeoDa version 1.4.0 (Anselin *et al.* 2006). The output of this smoothing was used to produce a relative risk map by calculating observed over expected proportions of positives for each polygon (Figure 4). The number used for expected proportion of positives was the per cent positive for the entire province (21.4% for total coliforms and 2.4% for *E. coli*). The output of this calculation was then used to detect areas of higher and lower risk using distance-weighted interpolation, which provided an interpolated relative risk for polygons where no data were available.

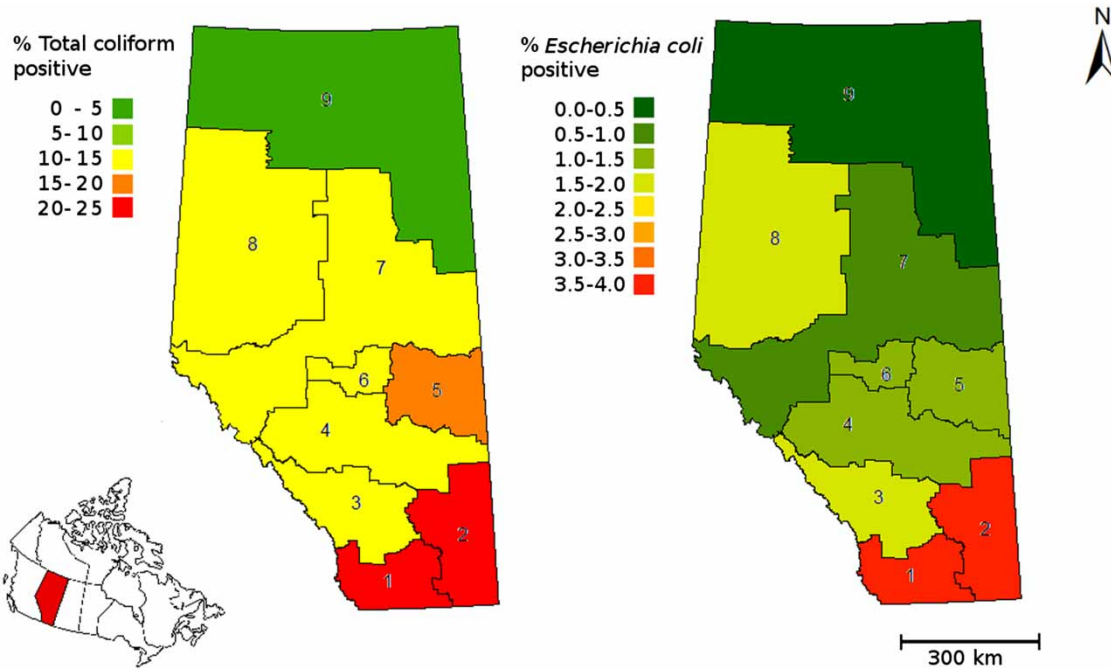


Figure 3 | *Escherichia coli* and total coliform per cent positivity (at the sampling level without aggregation by quarter section and month) by Alberta Health Services Health Region for private and public well water for the years 2004–2012, Alberta, Canada.

RESULTS

Geolocation

A survey method used in Western Canada, the ATS System, allows estimates of geographical coordinates. ATS data was provided by 72.6% ($n = 130,366$) of the water sample submitters. Tests having incomplete ATS data (5.8%, $n = 10,363$) as well as samples with complete but invalid ATS data (1.3%, $n = 2,406$) were excluded, resulting in 65.4% ($n = 117,597$) geolocated samples. In addition to geolocation, health region location of the water sample submitter was provided for 98.9% ($n = 177,618$) of the samples. The ATS coordinates allow for a parcel of land to be georeferenced to a resolution of 1 quarter section ($\sim 800 \times 800$ m or 0.65 km^2 (Alberta Environment and Parks 2010)).

Frequency of contamination, overall and by water source

Overall, including repeated samples, 14.6 and 1.5% of the well samples were total coliform- and *E. coli*-positive,

respectively. Prevalence of total coliform- and *E. coli*-positive wells in the data aggregated by quarter section and month was 21.4 and 2.4%, respectively.

Tests were evenly divided between private and public well water sources (49.9 and 50.1%, respectively). A larger proportion of private well water samples (81.1%, $n = 72,603$) was geolocated compared with public samples (50.0%, $n = 44,994$). The frequency of total coliform-positive wells in public vs. private wells was different between geolocated and non-geolocated wells (Breslow Day = 76.099, $df = 1$, $p < 0.001$), so geolocated and non-geolocated wells were presented separately. Private geolocated wells had 3.37 (95% CI: 3.24–3.51) higher odds of being total coliform-positive compared to public geolocated wells, while non-geolocated private wells had a 2.49 (95% CI: 2.35–2.63) higher odds compared to public non-geolocated wells. The difference between public and private was not different by geolocation status for *E. coli* contamination (Breslow Day = 1.233, $df = 1$, $p = 0.362$), so the CMH pooled odds ratio was used. A private well had 5.21 (95% CI: 4.66–5.82) higher odds of being *E. coli*-positive than a public well, irrespective of geolocation.

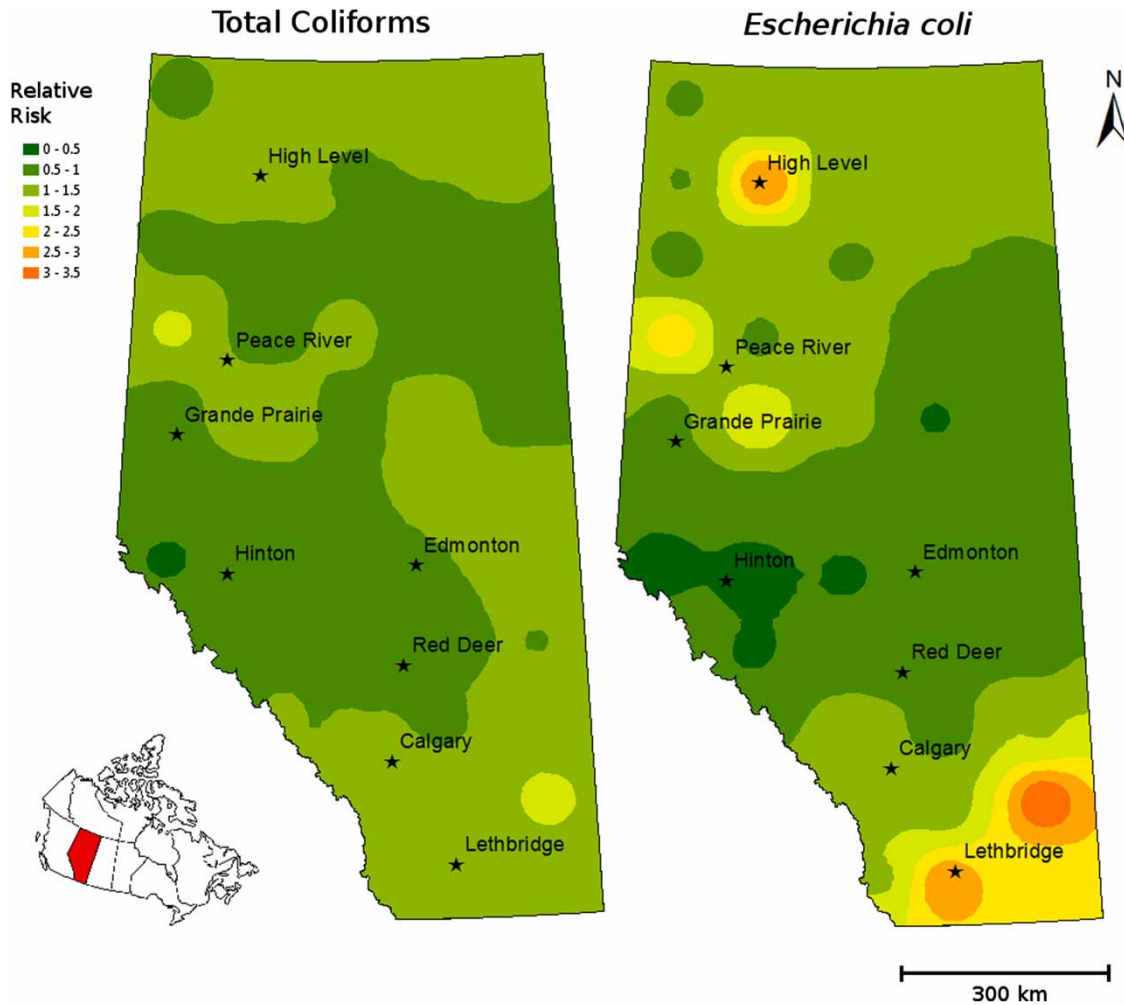


Figure 4 | Relative risk maps for Alberta for *Escherichia coli* and total coliforms calculated using data aggregated by quarter section and month for private and public well water, 2004–2012, Alberta, Canada.

Repeat testing

Public wells were more often repeatedly tested with a median of two times per year (Interquartile Range (IQR) 1,4) compared to private wells that were tested a median of one time per year (IQR 1,1) ($W = 126,112,944$, $p < 0.001$). The maximum number of repeats for a single quarter section per year for public wells was 399 and for private wells was 136.

Spatiotemporal analysis

The number of samples submitted for water quality testing varied over the week ($\chi^2 = 170,853$, $df = 6$, $p < 0.001$).

Samples were submitted most frequently on Tuesdays (29.5%) and Wednesdays (38.5%).

The STL decomposition demonstrated a peak in both *E. coli* and total coliform-positive wells in 2005, and a second lower spike for total coliforms in 2009 (Figures 1 and 2). The time series plot also demonstrated a peak in 2005 for both *E. coli* and total coliforms, but did not indicate a second peak for total coliforms. There was a seasonal spike in both *E. coli* ($\chi^2 = 1,224$, $df = 2$, $p < 0.001$) and total coliforms ($\chi^2 = 3,486$, $df = 2$, $p < 0.001$) in the summer and fall months, respectively, with peak dates for all years combined (2004–2012) of August 25 for total coliforms and July 24 for *E. coli* using Edward's test of seasonality.

Using the annual trends in the data aggregated by quarter section and month, the 2005 peaks for both total coliforms and *E. coli* were primarily limited to the southern region of Alberta. All three regions had significant seasonality by Edward's test for both total coliforms and *E. coli* (Table 1). Peak dates (averaged across all years) were similar for south and central regions of the province with *E. coli* and total coliforms peaks on July 23 and August 18 for south, and July 21 and September 6 for central Alberta, respectively. The northern regions of the province reflected a later seasonal pattern with peaks on August 6 and September 12, respectively.

Using all data (not aggregated by quarter section/month) and examining frequency maps for *E. coli* and total coliforms by health region, the maps revealed a south to north gradient with higher rates of both *E. coli*- and total coliform-positive wells in the southern part of the province, moderate rates in the central portion of the province and lower rates in the north (Figure 3).

Using the quarter section/month aggregated data, an area of high relative risk of *E. coli* contamination (Figure 4) was found in the south (ranging from 2.4 to 3.4) and three areas of higher relative risk were found in the north (one at 1.6–2.1, one at 1.7–2.3 and one at 1.6–3.2). The total coliforms relative risk map had a more uniform appearance with higher risk areas in the same locations as the *E. coli* relative risk map, but the risk levels were lower with a maximum relative risk of 1.6 (Figure 4).

DISCUSSION

Overall, use of these techniques to assess routine test data from across a broad region provides a basis for a framework for routine, passively collected surveillance data, giving insight about seasonality, temporality and geographical located relative risk. Use of the empirical Bayesian smoothing technique with the relative risk map allowed visualization of high-risk areas that were not seen with other methods.

The time series analyses provided a visual representation of the baseline levels of contamination and departures from the baseline and Edward's test provided a statistical means of testing seasonality. The seasonality in *E. coli*-positive wells corresponds with the seasonal trend

seen in human cases of *E. coli* O157:H7 (July peak) (Michel et al. 1999) and in prevalence of *E. coli* O157 in cattle rectally collected rectal faecal samples (peaks in spring and late summer) (Chapman et al. 1997). The STL decomposition demonstrates a peak in 2005 for *E. coli* and total coliform contamination. This corresponds with a flooding event that occurred in Alberta in 2005. Visually, the time series by region shows that the peak in 2005 was restricted to the southern part of the province, which is where the flooding occurred. We were unable to determine the cause of the second peak in 2009 in total coliform contamination revealed by the STL decomposition.

Contamination maps of *E. coli* and total coliforms aggregated by Alberta Health Services health region indicated areas of high contamination in the southern health regions 1 and 2, moderate rates in the centre of the province and low rates in the north (Figure 3). This is corroborated by earlier studies: spatial clusters of *E. coli* O157 cases in humans have been identified in the same general area of the province as one of the areas this study has identified as high risk for water contamination (Pearl et al. 2006) and incidence rates of human cases of cryptosporidiosis and campylobacteriosis are consistently higher for the south health zone than for Alberta overall (Government of Alberta 2015).

Relative risk maps created from point data aggregated by quarter section and month with empirical Bayesian smoothing applied identified the same areas of risk in the south of Alberta as the sample-level contamination maps (Figure 3), but also identified three areas of higher risk in the north-western part of the province (close to the cities Grand Prairie, Peace River and High Level; Figure 4). The elevated risk in these regions may be due to geographical features such as type of soil or aquifer, precipitation patterns, well depths or land use patterns, as well as socio-economic factors. Distance-weighted interpolation allowed us to interpolate relative risk to geographical regions without data. Relative risk maps will be useful as a guide for further investigations into identifying hot spots of microbial contamination in groundwater and as a geographic baseline for future surveillance.

Although this study uses data collected for other purposes, we suggest this form of surveillance with passively acquired data is generalizable to the larger population of wells across the province, understanding there may be

some bias in the sampling frame. In addition, using passively collected data increased the power of the study due to the large size of the database ($n = 179,623$). The overall per cent of *E. coli*- and total coliform-positive wells in this study, 1.5 and 14.6%, respectively, is lower than reported in studies from other regions of Canada and the United States (Goss *et al.* 1998; Borchardt *et al.* 2003; Hetcher-Aguila 2005). In Ontario, Canada, between 17 and 24% purposively selected wells on agricultural land were *E. coli*-positive at least once in a two-year period, with a higher prevalence of positives in summer (Goss *et al.* 1998). In Wisconsin, 28 and 2% of 50 wells sampled with a focus on septic field density were total coliform- or *E. coli*-positive in longitudinal study over one year (Borchardt *et al.* 2003). Of 24 untreated public supply wells and 13 private residential wells in the Chemung River Basin, New York, USA, sampled in the summer of 2003, 32% were total coliform-positive and 16% *E. coli*-positive. Site selection for this study was based on selecting sites with greater vulnerability to contamination as well as good representation of the geographical extent of the study (Hetcher-Aguila 2005). A total of 22% of UK private water sources, primarily groundwater, was positive for total coliforms, faecal coliforms or faecal streptococci (Fewtrell *et al.* 2007). The purposive nature of the sample selection in these studies alone, selecting for areas more likely to be contaminated (Goss *et al.* 1998; Borchardt *et al.* 2003; Hetcher-Aguila 2005), could account for the higher levels of contamination in other studies. However, lower levels of contamination in our study also may be due to bias in our sampling. Notably, most wells examined were voluntarily sampled by their owners and were not randomly selected from the population of wells at risk. A survey study in Alberta indicated that voluntary sampling of private wells represented only 11% of wells in service (Summers 2010). Specifically, bias would occur if people who provided samples were more conscientious in their well maintenance compared with people who did not provide samples. Alternatively, our sample would be biased in the opposite direction if people submitted water samples because of suspicions about water contamination. The impact of the voluntary bias could be better understood using a study which included both a survey portion to examine individual water testing habits coupled with a water test to indicate whether or not those inclined to test voluntarily are more

or less likely to have contamination. This would allow a better understanding of the feasibility of using voluntary samples in a surveillance system. However, similar to our findings, a study of 816 wells in 2001 in Alberta demonstrated a total coliform prevalence of 13.8% and a faecal coliform prevalence of 3.1% (no *E. coli*-specific testing was performed in this study, and the study design was based on convenience sampling with mandatory inclusion of sites in each of 64 municipalities) (Fitzgerald *et al.* 2001).

Unlike other study designs where sampling frequency is mandated, in our study, water samples from the same wells could be submitted more than once per year (repeated testing). Repeated testing was important to examine because contamination with both total coliforms and faecal coliforms can be sporadic (Oliphant *et al.* 2002). A positive test in a well that has only been tested once in a year may have different implications than a well that has been tested 100 times in a year with one positive finding. It is not possible from our data to determine if a positive test from wells that have only been tested once represents a sporadic contamination event, or a constant problem. As using the proportion of positive tests as the statistical unit would make a single positive test over the course of a year's worth of testing seem unimportant, we decided to aggregate to quarter section and month. This has the effect of potentially biasing the outcome towards higher positivity, but from a public health perspective, treating the positives under all conditions as serious is not unreasonable. In addition, the data may be misleading in that test submissions were aggregated by quarter section and multiple tests in a year may have been from one well tested multiple times or several wells tested one or more times. High numbers of tests for a single quarter section were likely occurring where there were multiple wells on the quarter section. It is impossible to calculate what bias this may have introduced into the results. Geolocation using the ATS system allowed us to use the large database that we had, but was not ideal. Global positioning system (GPS) coordinates would have allowed better resolution for spatial analysis, and better transferability to other regions.

A portion of the population may be at risk but is not testing their water on a regular basis or at all. Summers (2010) previously identified that few private well owners were testing annually, and identified the most frequent reason for not

testing was feeling there was no need to test. Regulated annual routine testing of private and small public system well water may be of benefit to Albertans, and our data supports the recommendations outlined in the *Guidance on Waterborne Bacterial Pathogens* by Health Canada (2013), where private and semi-public drinking water systems should be tested two or three times per year, with a focus on times when contamination is more likely, i.e., spring and summer.

Determining which samples were repeated testing was a difficult task on such a large database, as names were not entered consistently. For instance J. Smith and John Smith could represent the same person. A fuzzy lookup function (K2 Enterprises 2012) could potentially identify inconsistent name entries, but this approach would not identify cases where multiple samples from the same well were submitted by family members or friends with different names. To address this problem, we aggregated all geolocated samples in the same quarter section, and used quarter sections as our statistical unit.

The use of CMH allowed examination of the differences in contamination between public and private wells while controlling for geolocation. Although total coliform and *E. coli* contamination was higher for private wells than public wells, private well overseers tested their wells less frequently. The majority of small public water systems have no regulations for testing, with the exception of campgrounds, which require water testing just prior to opening in the spring (Province of Alberta 2004), but the guidelines suggest more frequent testing than for private systems, depending on the water source. For instance, it is recommended that treated surface water and treated groundwater under the influence of surface water be tested weekly for communal/public supplies (Technical Advisory Committee on Safe Drinking Water 2007). In addition, because of multiple people involved with small public water systems, more resources and organization may be in place to ensure testing, maintenance and disinfection. These factors may account for the lower contamination in public wells. A 2011 survey of private well owners in Alberta demonstrated most study participants had a lack of knowledge about groundwater sources and well management and only 11% tested for microbial contamination on at least a yearly basis (Summers 2010). As a result of the higher

contamination rates, private well owners do appear to be more at risk than public well users, so encouraging private well owners to test more frequently and providing more educational opportunities for private well owners may be of value.

Limitations of this study include the previously mentioned, largely voluntary nature of the test samples, allowing inference only to this population of tested well water samples, and the inability to geolocate all samples. In addition, private samples were better represented in the relative risk maps because it was possible to geolocate more private samples (81.1%, $n = 72,603$) than public samples (50.0%, $n = 44,994$), and this increased the contamination rates as private well water samples had higher contamination than public well water samples. The ability to geolocate all the samples may have changed the map and, potentially, the location of higher risk areas.

CONCLUSIONS

Using geolocation with empirical Bayes smoothing with a substantial, passively collected database of water tests, we were able to successfully identify new areas of concern in addition to corroboration of previously identified hot spots of contamination. This kind of information can be used by decision-makers to create targeted surveillance in these high-risk areas. Coordinates such as the ATS system used in this study are not ideal. Passively collected data are strengthened by the addition of GPS coordinates. Although passively collected data have limitations, they are an economical way to perform some types of surveillance and provide baseline information on contamination levels and trends over several years.

ACKNOWLEDGEMENTS

This study was funded by the Government of Alberta, the University of Calgary, Department of Ecosystem and Public Health, the Natural Sciences and Engineering Research Council of Canada, and Mitacs. We appreciate the contribution of Iqbal Jamal and his company, AQL Management, to the Mitacs funding. The authors wish to

acknowledge the contributions of the personnel at ProvLab, especially Jennifer May-Hadford, Sumana Fathima and Lorraine Ingham for their help with data, maps and laboratory protocols. In addition, we appreciate Dr Henrik Stryhn's statistical advice and help provided by Peter Peller of the University of Calgary Library with geographic information data and software.

REFERENCES

- Abramson, J. H. 2011 WINPEPI updated: computer programs for epidemiologists, and their teaching potentials. *Epidemiol. Perspect. Innov.* **8** (1), 1–9.
- Alberta Environment 2009 Alberta environment's drinking water program: A 'Source to Tap, Multi-Barrier' approach. <http://aep.alberta.ca/water/programs-and-services/drinking-water/documents/DrinkingWaterProgram-May2009.pdf> (accessed 28 March 2015).
- Alberta Environment and Parks 2010 Alberta Township Survey System. Alberta Environment and Sustainable Resource Development. <http://esrd.alberta.ca/recreation-public-use/recreation-on-agricultural-public-land/alberta-township-survey-system.aspx> (accessed 16 June 2013).
- Anselin, L., Ibnu, S. & Youngihn, K. 2006 Geoda: an introduction to spatial data analysis. *Geogr. Anal.* **38**, 5–22. doi:10.1111/j.0016-7363.2005.00671.x.
- Anselin, L., Kim, Y. W. & Syabri, I. 2010 Web-based analytical tools for the exploration of spatial data. In: *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* (M. Fischer & A. Getis, eds). Springer, Heidelberg, Germany.
- Bezirtozoglou, C., Dekas, K. & Charvalos, E. 2011 Climate changes, environment and infection: facts, scenarios and growing awareness from the public health community within Europe. *Anaerobe* **17**, 337–340. doi:10.1016/j.anaerobe.2011.05.016.
- Borchardt, M., Bertz, P., Spencer, S. & Battigelli, D. 2003 Incidence of enteric viruses in groundwater from household wells in Wisconsin. *Appl. Environ. Microbiol.* **69** (2), 1172–1180. doi:http://dx.doi.org/10.1128/AEM.69.2.1172-1180.2003.
- Chapman, P. A., Siddons, C. A., Cerdan Malo, A. T. & Harkin, M. A. 1997 A 1-year study of *Escherichia coli* o157 in cattle, sheep, pigs and poultry. *Epidemiol. Infect.* **119**, 245–250. doi:http://dx.doi.org.ezproxy.lib.ucalgary.ca/10.1017/S0950268897007826.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. 1990 STL: a seasonal-trend decomposition procedure based on loess. *J. Off. Stat.* **6**, 3–73.
- Cook, C., Prystajeky, N., Feze, I. N., Joly, Y., Dunn, G., Kirby, E., Özdemir, V. & Isaac-Renton, J. 2013 A comparison of the regulatory frameworks governing microbial testing of drinking water in three Canadian provinces. *Can. Water Resour. J. Rev. Can. Resour. Hydr.* **38**, 185–195. doi:10.1080/07011784.2013.822186.
- Craighead, L., Gilbert, W., Subasinghe, D. & Häslér, B. 2015 Reconciling surveillance systems with limited resources: an evaluation of passive surveillance for rabies in an endemic setting. *Prev. Vet. Med.* **121**, 206–214. doi:10.1016/j.prevetmed.2015.06.016.
- Cressie, N. 1995 Bayesian smoothing of rates in small geographic areas. *J. Reg. Sci.* **35**, 659–673. doi:10.1111/j.1467-9787.1995.tb01298.x.
- Federal-Provincial-Territorial Committee on Drinking Water, CCME Water Quality Task Group 2004 From source to tap: Guidance on the multi-barrier approach to safe drinking water. <http://www.hc-sc.gc.ca/ewh-semt/water-eau/drink-potab/multi-barrier/index-eng.php> (accessed 28 March 2014).
- Fewtrell, L., Kay, D. & Godfree, A. 2007 The microbiological quality of private water supplies. *Water Environ. J.* **12**, 45–47. doi:10.1111/j.1747-6593.1998.tb00145.x.
- Fidalgo, Á. M. & Madeira, J. M. 2008 Generalized Mantel-Haenszel methods for differential item functioning detection. *Educ. Psychol. Meas.* **68**, 940–958. doi:10.1177/0013164408315265.
- Fitzgerald, D., Chanasyk, D. S., Neilson, R. D., Kiely, D. & Audette, R. 2001 Farm well water quality in Alberta. *Water Qual. Res. J. Can.* **36**, 565–588.
- Galanis, E., Mak, S., Otterstatter, M., Taylor, M., Zubel, M., Takaro, T. K., Kuo, M. & Michel, P. 2014 The association between campylobacteriosis, agriculture and drinking water: a case-case study in a region of British Columbia, Canada, 2005–2009. *Epidemiol. Infect.* **142**, 2075–2084. doi:10.1017/S095026881400123X.
- Galway, L. P., Allen, D. M., Parkes, M. W., Li, L. & Takaro, T. K. 2015 Hydroclimatic variables and acute gastro-intestinal illness in British Columbia, Canada; a time series analysis. *Water Resour. Res.* **51**, 885–895. doi:10.1002/2014WR015519.
- Gleeson, C. & Gray, N. 1996 *The Coliform Index and Waterborne Disease: Problems of Microbial Drinking Water Assessment*. E & FN Spon, London, UK.
- Goss, M., Barry, D. A. & Rudolph, D. 1998 Contamination in Ontario farmstead domestic wells and its association with agriculture: 1. Results from drinking water wells. *J. Contam. Hydrol.* **32** (3–4), 267–293. doi:http://dx.doi.org/10.1016/S0169-7722(98)00054-0.
- Government of Alberta 2015 Notifiable Diseases – Age-Sex Specific Incidence Rate (common diseases). http://www.ahw.gov.ab.ca/IHDA_Retrieval/selectResults.do (accessed 6 June 2015).
- Hafen, R. P., Anderson, D. E., Cleveland, W. S., Maciejewski, R., Ebert, D. S., Abusalah, A., Yakout, M., Ouzzani, M. & Grannis, S. J. 2009 Syndromic surveillance: STL for modeling, visualizing, and monitoring disease counts. *BMC Med. Inform. Decis. Mak.* **9**, 1–11. doi:10.1186/1472-6947-9-21.

- Hay, S. I., George, D. B., Moyes, C. L. & Brownstein, J. S. 2013 Big data opportunities for global infectious disease surveillance. *PLoS Med.* **10**, 1–4. doi:10.1371/journal.pmed.1001413.
- Health Canada 2013 *Guidance on waterborne bacterial pathogens*. http://publications.gc.ca/collections/collection_2014/sc-hc/H129-25-1-2014-eng.pdf (accessed 16 June 2015).
- Hetcher-Aguila, K. K. 2005 *Ground-Water Quality in the Chemung River Basin, New York, 2003 (No. 01961497)*. US Geological Survey, Reston, VA, USA.
- IDEXX Laboratories 2013 Colilert. <http://www.idexx.com/resource-library/water/colilert-procedure-en.pdf> (accessed 8 October 2013).
- K2 Enterprises 2012 Performing fuzzy lookups in Excel. Continuing Professional Education for Accounting and Financial Professionals. <http://www.k2e.com/tech-update/tips/431-tip-fuzzy-lookups-in-excel> (accessed 26 March 2013).
- Krolik, J., Maier, A., Evans, G., Belanger, P., Hall, G., Joyce, A. & Majury, A. 2013 A spatial analysis of private well water *Escherichia coli* contamination in southern Ontario. *Geospatial Health* **8**, 65–75. doi:http://dx.doi.org/10.4081/gh.2013.55.
- Michel, P., Wilson, J. B., Martin, S. W., Clarke, R. C., McEwen, S. A. & Gyles, C. L. 1999 Temporal and geographical distributions of reported cases of *Escherichia coli* o157:H7 infection in Ontario. *Epidemiol. Infect.* **122**, 193–200. doi:10.1017/S0950268899002083.
- Oliphant, J., Ryan, M., Chu, A. & Lambert, T. 2002 Efficacy of annual bacteria monitoring and shock chlorination in wells finished in a floodplain aquifer. *Ground Water Monit. Remediat.* **22** (4), 66–72. doi:http://dx.doi.org/10.1111/j.1745-6592.2002.tb00772.x.
- Owusu-Edusei Jr, K. & Owens, C. J. 2009 Monitoring county-level chlamydia incidence in Texas, 2004–2005: application of empirical Bayesian smoothing and exploratory spatial data analysis (ESDA) methods. *Int. J. Health Geogr.* **8**, 1–6. doi:10.1186/1476-072X-8-12.
- Pardhan-Ali, A., Berke, O., Wilson, J., Edge, V. L., Furgal, C., Reid-Smith, R., Santos, M. & McEwen, S. A. 2012 A spatial and temporal analysis of notifiable gastrointestinal illness in the Northwest Territories, Canada, 1991–2008. *Int. J. Health Geogr.* **11**, 17–26. doi:10.1186/1476-072X-11-17.
- Pearl, D. L., Louie, M., Chui, L., Doré, K., Grimsrud, K. M., Leedell, D., Martin, S. W., Michel, P., Svenson, L. W. & McEwen, S. A. 2006 The use of outbreak information in the interpretation of clustering of reported cases of *Escherichia coli* o157 in space and time in Alberta, Canada, 2000–2002. *Epidemiol. Infect.* **134**, 699–711. doi:10.1017/S0950268805005741.
- Province of Alberta 2004 Public Health Act: Recreation area regulation: Alberta Regulation 198/2004 with amendments up to and including Alberta Regulation 85/2012. http://www.qp.alberta.ca/1266.cfm?page=2004_198.cfm&leg_type=Regs&isbncln=9780779765119&display=html (accessed 7 May 2016).
- Samadder, S. R., Ziegler, P., Murphy, T. M. & Holden, N. M. 2010 Spatial distribution of risk factors for *Cryptosporidium* spp. transport in an Irish catchment. *Water Environ. Res.* **82**, 750–758. doi:http://dx.doi.org/10.2175/106143010x12609736966649.
- Shumway, R. & Stoffer, D. 2006 *Time Series Analysis and Its Applications: With R Examples*. Springer, New York, NY, USA.
- Simpson, M. W. M., Allen, D. M. & Journeay, M. M. 2014 Assessing risk to groundwater quality using an integrated risk framework. *Environ. Earth Sci.* **71**, 4939–4956. doi:10.1007/s12665-013-2886-x.
- Standridge, J. 2008 *E. coli* as a public health indicator of drinking water quality. *J. Am. Water Works Assoc.* **100**, 65–75.
- Statistics Canada 2014 Population by sex and age group, by province and territory (Number, both sexes). <http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo31a-eng.htm> (accessed 6 November 2014).
- Stevenson, M., Stevens, K. B., Rogers, D. J. & Clements, A. C. A. 2008 *Spatial Analysis in Epidemiology*, 1 edn. Oxford University Press, New York, USA.
- Summers, R. 2010 Alberta water well survey: A report prepared for Alberta Environment. <http://environment.gov.ab.ca/info/library/8337.pdf> (accessed 7 May 2016).
- Technical Advisory Committee on Safe Drinking Water 2004 Environmental Public Health Field Manual for Private, Public and Communal Drinking Water Systems in Alberta, 2nd edn. <http://www.health.alberta.ca/documents/Drinking-Water-Systems-2004.pdf> (accessed 7 May 2016).
- Thrusfield, M., Noordhuizen, J. P., Frankena, K., Ortega, C. & de Blas, I. 2001 WIN EPISCOPE 2.0: improved epidemiological software for veterinary medicine. *Vet. Rec. J. Br. Vet. Assoc.* **148**, 567–572. doi:http://dx.doi.org/10.1136/vr.148.18.567.
- United States Environmental Protection Agency 2012 Private Drinking Water Wells. <http://water.epa.gov/drink/info/well/index.cfm> (accessed 19 October 2015).
- United States Environmental Protection Agency 2015 Public Drinking Water Systems Programs. <http://water.epa.gov/infrastructure/drinkingwater/pws/index.cfm> (accessed 19 October 2015).
- Vijay, R., Ramya, S. S., Pujari, P. R. & Mohapatra, P. K. 2011 Spatio-temporal assessment of ground water level and quality in urban coastal city Puri, India. *Water Sci. Technol. Water Supply* **11**, 194–201. doi:10.2166/ws.2011.021.
- World Health Organization 2011 *Guidelines for Drinking-Water Quality*, 4th edn. WHO, Geneva, Switzerland. http://www.who.int/water_sanitation_health/publications/2011/dwq_chapters/en/ (accessed 20 January 2014).

First received 11 May 2016; accepted in revised form 28 May 2017. Available online 13 July 2017