

# A spatial analysis of private well water *Escherichia coli* contamination in southern Ontario

Julia Krolik<sup>1,2</sup>, Allison Maier<sup>1</sup>, Gerald Evans<sup>2,3,4</sup>, Paul Belanger<sup>5,6,7</sup>, Geoffrey Hall<sup>8</sup>, Alan Joyce<sup>1</sup>, Anna Majury<sup>1,2,4,7</sup>

<sup>1</sup>Public Health Ontario, Regional Ontario, Canada; <sup>2</sup>Department of Pathology and Molecular Medicine, Queen's University, Kingston, Ontario, Canada; <sup>3</sup>Division of Infectious Diseases, Department of Medicine, Queen's University and Kingston General Hospital, Kingston, Ontario, Canada; <sup>4</sup>Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario, Canada; <sup>5</sup>Department of Geography, Queen's University, Kingston, Ontario Canada; <sup>6</sup>Kingston Frontenac Lennox and Addington Health Unit, Kingston, Ontario, Canada; <sup>7</sup>Department of Public Health Sciences, Queen's University, Kingston, Ontario, Canada; <sup>8</sup>Department of Civil Engineering, Queen's University, Kingston, Ontario, Canada

**Abstract.** Research to date has provided limited insight into the complexity of water-borne pathogen transmission. Private well water supplies have been identified as a significant pathway in infectious disease transmission in both the industrialised and the developing world. Using over 90,000 private well water submission records representing approximately 30,000 unique well locations in south-eastern Ontario, Canada, a spatial analysis was performed in order to delineate clusters with elevated risk of *E. coli* contamination using 5 years of data (2008-2012). Analyses were performed for all years independently and subsequently compared to each other. Numerous statistically significant clusters were identified and both geographic stability and variation over time were examined. Through the identification of spatial and temporal patterns, this study provides the basis for future investigations into the underlying causes of bacterial groundwater contamination, while identifying geographic regions that merit particular attention to public health interventions and improvement of water quality.

**Keywords:** spatial cluster analysis, private well water, *Escherichia coli* contamination, public health, Canada.

---

## Introduction

Any compromise to drinking water integrity poses a serious threat to human, animal and environmental health and is a major health issue in developed nations to this day. Water quality is jeopardised by microbiological and/or chemical agents, water-borne bacteria, viruses and protozoa posing the greatest threat (Charrois, 2010). In North America, contaminated drinking water is conservatively estimated to cause one thousand deaths and one million illnesses per year (Sierra Legal Defense Fund, 2006) and Europe shows similar rates (Charrois, 2010). Thus, safety of drinking water is a fundamental public health priority, with drinking water quality management being a crucial component in the prevention and control of water-borne disease (WHO, 2010).

Specifically, faecal contamination of well water is a serious public health issue, which has been linked to outbreaks of various water-borne infections (Furtado et al., 1998; Raina et al., 1999; Macler and Merkle, 2000; Corkal et al., 2004). There were 288 confirmed documented outbreaks of infectious enteric diseases in Canadian drinking water with the most common pathogens being *Giardia intestinalis*, *Campylobacter* sp., *Salmonella* and rotavirus over a 27-year period (Schuster et al., 2005). In 2000, the municipal water supply in Walkerton, Ontario was contaminated with *Escherichia coli* O157 from manure runoff originating in a nearby farm. It is to date, the largest municipal water-borne outbreak of *E. coli* O157 in Canadian history and it resulted in at least seven deaths and 2,300 cases of illness (Salvadori et al., 2009). Since the outbreak in Walkerton, numerous policies and public health actions have been implemented to prevent future outbreaks in municipal water supplies (Ontario Ministry of the Environment, 2009). However, the quality and maintenance of private well water systems remains the responsibility of the owner (Kreutzweiser et al.,

---

Corresponding author:  
Anna Majury  
Public Health Ontario Laboratories  
181 Barrie Street, Kingston, ON K7L 4V6, Canada  
Tel. +1 613 548 6630; Fax +1 613 547 1185  
E-mail: anna.majury@oahpp.ca

2011). In Canada, approximately three to four million, or 1 in 10 persons rely on private well water sources (Statistics Canada, 2011), which remain vulnerable to contamination and should be properly monitored (Charrois, 2010). In England and Wales, the incidence rate of intestinal infectious outbreaks in recipients of private water supplies was 35 times that of those receiving public water supplies (Smith et al., 2006). Within Canada, a study in British Columbia found that the risk of enteric disease was 5.2 times higher for individuals living on land serviced by private wells than those who relied upon municipal groundwater systems (Uhlmann et al., 2009).

Understanding of the endemic water-borne risks associated with private wells and groundwater is currently lacking (Uhlmann et al., 2009). Surveillance of contaminated wells and identification of clusters with a high risk of contamination would allow public health practitioners to better assess and design interventions and aid in the identification of contributing factors leading to contamination. This can be achieved using geographical information systems (GIS) and other spatial analytical tools. For example, spatial clustering can be employed to detect foci of well water contamination. Although many studies have employed the spatial scan statistic to detect clusters of various pathogens in specific geographic locations (Brownstein et al., 2002; Odoi et al., 2004; Coleman et al., 2009; Szonyi et al., 2010), few have used this method to investigate clusters pertaining to water-borne infections or water contamination in Canada. *E. coli* is the principal bacteriological indicator used in Ontario to assess faecal contamination of drinking water (Ontario Ministry of the Environment, 1994) and Public Health Ontario (PHO) provides bacteriological testing of private well water samples (as a free service). A previous study of 235 rural wells in Ontario found that 9.5% of households had at least one *E. coli* - positive water sample (Strauss et al., 2001).

As the first study to conduct spatial analysis of bacteriological contamination in private well water, we aimed to provide a geospatial description of the prevalence and risk of *E. coli* contamination in south-eastern Ontario for a 5-year period (2008-2012). Furthermore, possible geographical trends were explored on a temporal scale and, as private well water contamination is widespread, cluster analysis was performed at the provincial level for 2012 to investigate other regions with elevated risk of *E. coli* contamination.

## Material and methods

### *Data preparation and GIS*

Private well water quality data was compiled from ongoing bacteriological testing at Public Health Ontario Laboratories (PHOLs), resulting in the use of a convenience sample (i.e. routinely collected data that is subsequently used for data analysis not intended at collection). In this case, the submitter is responsible for collection and transportation prior to bacteriological testing. All such data are stored in the Water Testing Information System (WTIS) database owned by PHO. For the present study, all records from samples submitted to the Kingston PHOL (PHOL-K) between January 1, 2008 and December 31, 2012 were retrieved from the database. A 5-year study period was chosen to limit spatial sampling bias, as the convenience sample differed by year. Records that were rejected due to improper sample handling were not included. The resulting dataset contained 107,547 records, however 3,723 had insufficient address information for geospatial analysis and were removed. The remaining records were geocoded as summarised in Fig. 1 using Google Earth™ (Google Inc., Google Earth™ version 5.1, <http://www.google.com/earth/index.html>) to obtain geocoordinates (geocodes). Google Earth™ only allows input of street address, city and postal code and thus, for the primary stage, city was defined as “city/town/municipality”. A secondary stage was performed utilising “county” for the city field. Given the rural nature of the wells, the provided information consisted of hamlets too small to be recognised by Google Earth™ or outdated place names resulting from boundary redefinitions or municipal amalgamations. Thus, geocoding was difficult and the two-step, rather than single-step, Google Earth™ search allowed for an additional 10% of records to be included in the study. This two-step process was then repeated on any non-geocoded records using LIOcoder, a service provided by the Ontario Ministry of Natural Resources (<http://www.lio.ontario.ca/liocoder/index.jsp>) and then ArcGIS version 10.0 (North America Geocode Service, CAN\_RoofTop locator, 2010) (ESRI Inc.; Redlands, USA). To ensure similarity in the geocoordinates provided by the different systems, a pilot comparison of geocoordinates was performed (data not shown). Between each geocoding software stage, manual geocoding was done by matching multiple submissions from the same address. Due to address quality (e.g. spelling or data entry mistakes or incorrect information), the automated systems did not always assign all

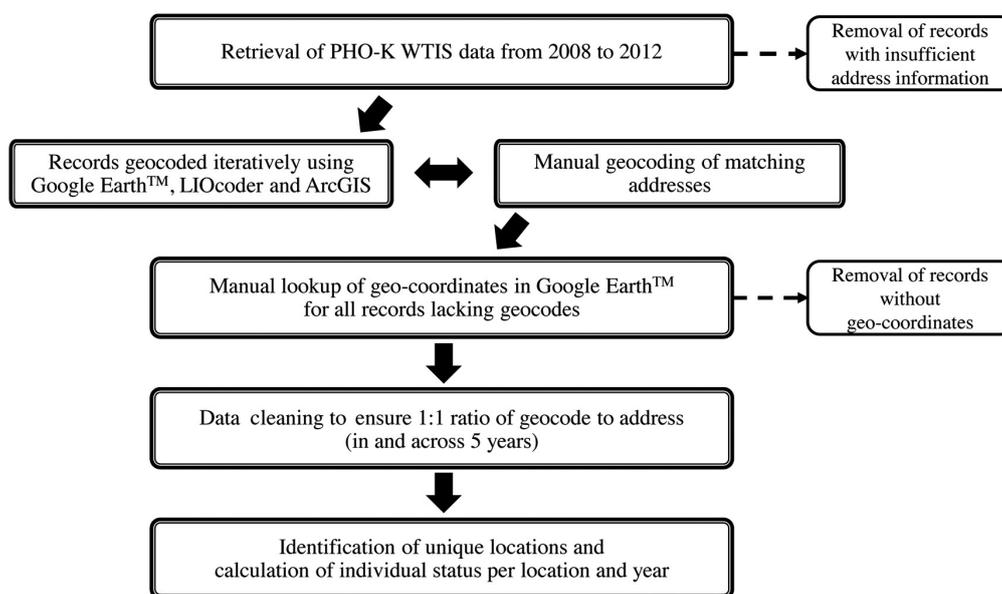


Fig. 1. Summary of the geocoding process.

records from the same address a geocode. Manual matching ensured that once a geocode was identified for an address, all submissions from the corresponding address were given the appropriate geocode. After all stages of automatic geocoding and manual matching were completed, geocoordinates for the remaining non-geocoded records were further determined manually using Google Earth™.

Once the geocoding process was complete, data cleaning methods were undertaken to ensure an address and a set of geo-coordinates had a one-to-one ratio. In some instances, a single geocode was assigned to more than one unique address and these records were removed. There were also cases where a single address was assigned multiple geocoordinates; a custom algorithm was applied to these to determine the most appropriate geocode. This data cleaning was performed first on individual years and then across all years of data, leaving a dataset of 90,149 geocoded well water records. As such, geocoordinates were found for 83.8% of the original 5-year dataset (between 81.0% and 85.6% by year). Table 1 outlines the number of records yearly and overall in the initial and final geocoded datasets, as well as their *E. coli* positive rates. As shown in Table 1, between 9,404 and 10,396 unique well locations were identified by year; when combined this resulted in 30,687 unique locations across all years. Greater than 98% of all unique locations were within four public health units (PHUs): Hastings Prince Edward County (HPEC),

Kingston Frontenac Lennox & Addington (KFL&A), Leeds Grenville Lanark (LGL) and Eastern Ontario (Fig. 2). For spatial analysis, *E. coli* status was determined yearly for unique locations; if a well had an *E. coli* positive submission in any given year it was designated as an *E. coli* contaminated well.

Table 1. Well water submissions from 2008 to 2012 to PHO-K: geocoding and *E. coli* summary results.

Year	Category	Number of records	<i>E. coli</i> - positive percentage
2008	All	23,744	6.34%
	Geocoded	20,326	7.20%
	Unique locations	10,396	9.09%
2009	All	22,034	5.99%
	Geocoded	18,096	5.91%
	Unique locations	9,574	7.98%
2010	All	20,572	6.35%
	Geocoded	17,535	6.35%
	Unique locations	9,706	8.47%
2011	All	20,081	7.14%
	Geocoded	17,078	6.38%
	Unique locations	9,404	8.71%
2012	All	21,116	4.42%
	Geocoded	17,114	4.63%
	Unique locations	9,727	6.25%
All years	All	107,547	6.04%
	Geocoded	90,149	5.95%
	Unique locations	30,687	8.11%

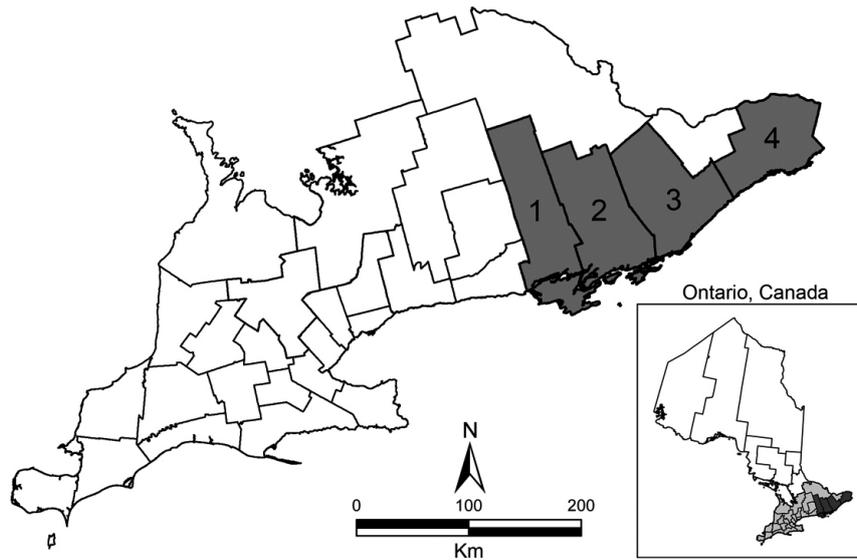


Fig. 2. Study region (both southern and south-eastern Ontario). Dark grey - south-eastern Ontario. white - rest of southern Ontario. 1 - HPEC; 2 - KFL&A; 3 - LGL; and 4 - Eastern Ontario.

In addition to the 2008-2012 dataset of private well water submissions to the PHOL-K, a second dataset consisting of 2012 well submissions to all the PHOLs in the province was retrieved. Only one year was retrieved, as it was not feasible to geocode 5 years of provincial data. The first round of geocoding was performed (two-step Google Earth™ and manual geocoding of matching addresses) which resulted in 118,503 records geocoded out of an original 185,045 (64.0%). The records represented 51,948 unique locations across the province. However, there was a differential level of geocoding between the PHOLs in the northern portion of the province relative to the southern (60.1% to 70.3% for the southern PHOLs compared to 33.2% to 59.1% for the northern PHOLs). Focusing on the southern portion to ensure sufficient levels of geocoding were met, the expanded geographic region was defined as southern Ontario (Fig. 2). The final dataset contained 47,048 unique locations (90.6% of all unique locations), which were assigned an *E. coli* status using the same definitions as the 5-year dataset. 5.21% of unique locations in southern Ontario had a least one *E. coli* positive result in 2012.

#### *Spatial clustering*

SaTScan™ software was used to perform spatial analysis on multiple datasets (M. Kuldorff and Information Management Service, Inc. SaTScan™ version 9.1.1; [www.satscan.org](http://www.satscan.org), 2013). This spatial

scan statistic investigates the occurrence of clusters by using a circular window of variable radius that systematically moves across the map (Kuldorff, 1997). The radius increases from zero to a user-defined maximum limit. This type of spatial analysis identifies clusters by comparing observed cases within the current radius of the window to the number of expected cases, provided they are randomly distributed. The location and statistical significance (P-values) were determined by carrying out 999 Monte Carlo replications, using the Bernoulli distribution (Kuldorff, 1997). A P-value less, or equal to 0.05, was used for primary clusters for the rejection of the null hypothesis of no clusters; and a P-value of 0.06 was used as the cut-off for secondary clusters, as the P-values generated for secondary clusters are higher. This spatial analysis was performed on datasets for each of the 5 years in the study (2008 to 2012) to identify clusters of *E. coli* contamination of private wells. Following the individual year analyses, the results were compared to observe temporal trends and variations. For spatial analysis, two different radii of the circular window were used (50% and 5% maximum population). Spatial analysis was also performed to identify clusters on all *E. coli* positives for the 2012 southern Ontario dataset, but only using 50% radii. All visualization of spatial results and subsequent mapping was done using ArcMap version 10.0. To better visualize the overlapping nature of the 5 year clusters, inner points were removed to clearly compare between clusters.

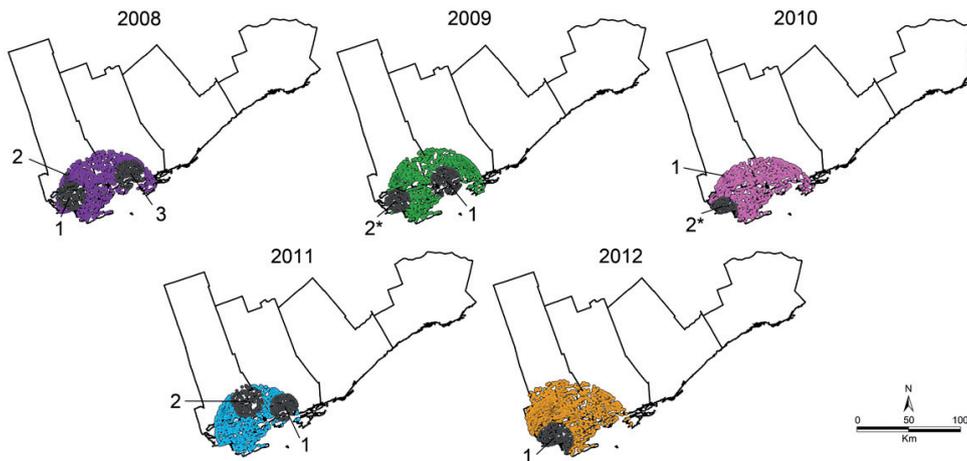


Fig. 3. Maps of 50% and 5% clusters by year. Purple - 2008; Green - 2009; Pink - 2010; Blue 2011; and Yellow - 2012. Colour represents 50% cluster, while grey represents 5% cluster(s). Numbers mark multiple individual 5% clusters and asterisks denotes secondary 5% cluster.

**Results**

The analysis identified a single statistically significant cluster for each of the 5 years when using a maximum population size of 50% within the cluster. These clusters can be observed in Fig. 3 (independently) and Fig. 4 (overlapped) and the corresponding P-values, relative risks (RR) and log likelihood ratios (LLR) can be found in Table 2. All of the clusters were in the same general region (the southern portion of HPEC and KFL&A PHUs), and the differences among years are visible in Fig. 4. The relative risks for the 5 years are consistent with a minimum of 1.48 and a maximum of 1.89.

A total of 10 clusters using the maximum of 5% of the population parameter were identified among the five years, with between one to three clusters per year.

Figs. 3 and 4 show these clusters independently by year and overlapped, respectively. Table 2 contains the corresponding P-values and RR. The colour and cluster numbering system is consistent across the figures and table. Two of the clusters (one from 2009 and one from 2010) were secondary clusters with P-values above the standard 0.05 but below the secondary P-value cut-off of 0.06.

There were four distinct geographical regions (henceforth designated by Roman numerals) with significant clusters: (i) the southwestern corner of HPEC (existing for 4 years); (ii) a small cluster to the north of the larger HPEC cluster (existing for 2 years); (iii) a cluster in the south central-west of KFL&A (existing for 3 years); and (iv) one slightly north of (iii) on the border between HPEC and KFL&A (existing for one year). The specific variations in location and size can

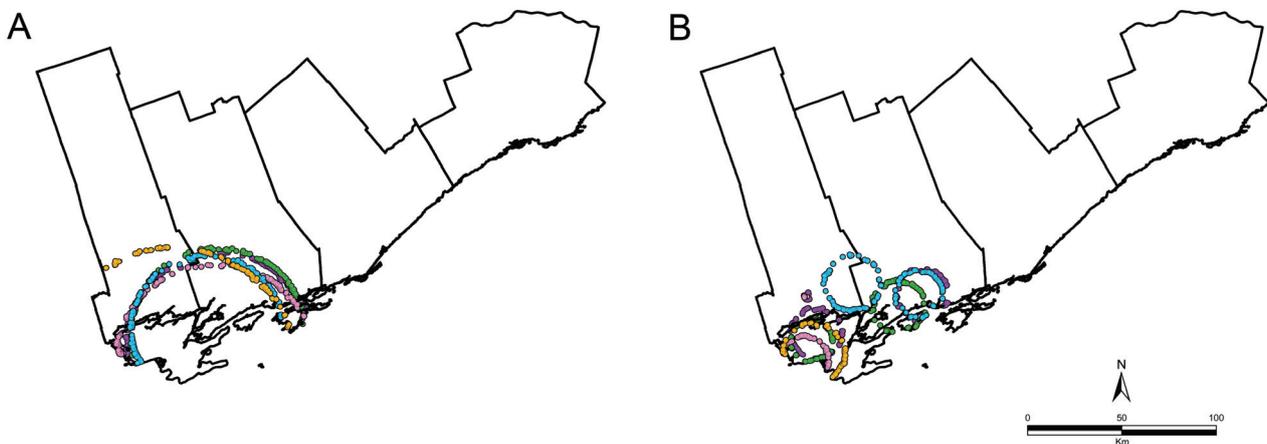


Fig. 4. Map of clusters overlapped from 2008 to 2012 at 5% and 50%. A - 50% clusters and B - 5% clusters. Purple - 2008; Green - 2009; Pink - 2010; Blue - 2011; Yellow - 2012.

Table 2. Statistical and geographical results of spatial analysis on individual years at 50% and 5% maximum population size.

Year	50% cluster						Cluster no.	5% cluster(s)					
	P-value	RR	LLR	Latitude (N)	Longitude (W)	Radius (km)		P-value	RR	LLR	Latitude (N)	Longitude (W)	Radius (km)
2008	<0.01	1.86	48.30	44.032760	76.909154	49.12	1	<0.01	2.07	20.61	44.076665	77.348221	13.90
							2	<0.01	3.39	16.16	44.253340	77.414734	2.09
							3	<0.01	1.94	16.04	44.269917	76.657330	14.45
2009	<0.01	1.77	32.97	44.076031	76.912635	46.79	1	<0.01	2.41	25.27	44.211622	76.812897	14.61
							2*	0.051	2.11	11.63	44.037839	77.375707	11.71
2010	<0.01	1.48	15.57	43.933029	76.923211	54.02	1	0.01	9.22	12.69	44.247490	77.435310	0.26
							2*	0.052	2.19	12.02	43.945648	77.422858	13.14
2011	<0.01	1.81	35.16	44.094944	76.985503	40.94	1	<0.01	2.01	13.88	44.273199	76.678298	13.17
							2	0.01	2.03	13.07	44.320444	77.119930	15.36
2012	<0.01	1.89	32.50	43.878538	77.219574	68.58	1	<0.01	2.56	23.36	43.976624	77.359530	16.85

\*Secondary cluster

be seen in the overlapped map in Fig. 4. For cluster (i) there was minimal variation in RR across the 4 years (2.07 to 2.56). These RRs were similar to those of cluster (iii) and (iv). Cluster (ii) had the highest RR (and smallest size), but there was high variation in risk (3.39 *versus* 9.22).

Six clusters were identified by spatial analysis of the Southern Ontario dataset using a maximum 50% radius; with one of these clusters being a secondary cluster. These clusters are represented on a map of southern Ontario in Fig. 5 with Table 3 containing their respective P-values, RRs and LLRs. The first cluster is in the same region as the 50% clusters identified by the regional analyses. The RR for this cluster was slightly higher than those found previously.

Clusters 2 and 3 are both large geographic regions (Niagara and Bruce Peninsula, respectively), with cluster 4 being geographically close to cluster 3.

## Discussion

There are health implications with consuming compromised water from rural sources (Raina et al., 1999; Said et al., 2003; Uhlmann et al., 2009; Charrois, 2010). Previous studies have focused on chemical contaminants such as nitrate, arsenic and pesticides (Goss et al., 1998; Rudolph et al., 1998; Benson et al., 2006; Knobloch et al., 2013) or bacteriological quality without spatial consideration (Goss et al., 1998; Rudolph et al., 1998; Bacci and Chapman, 2011;

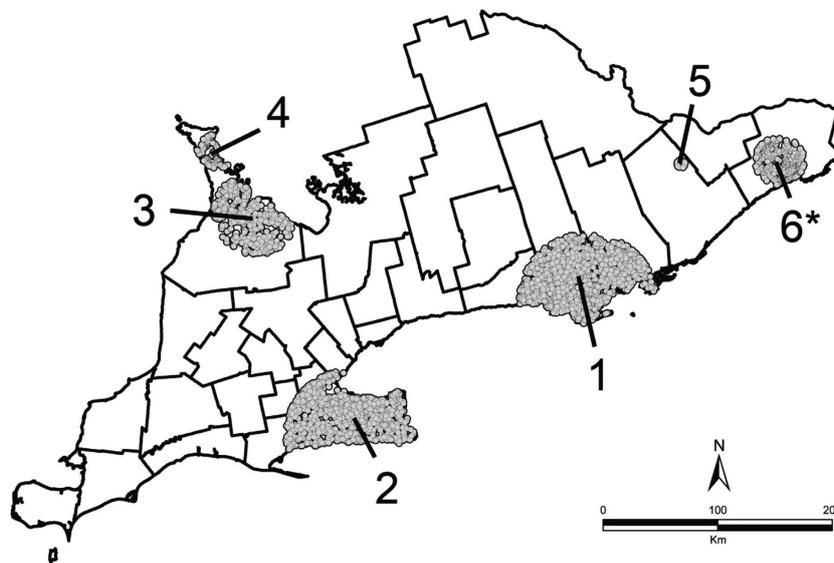


Fig. 5. Southern Ontario map of clusters at 50% maximum population size. Numbers mark individual clusters.

Table 3. Southern Ontario spatial analysis statistical and geographical cluster results (at 50% maximum population).

Cluster no.	P-value	RR	LLR	Latitude (N)	Longitude (W)	Radius (km)
1	<0.01	2.07	75.05	44.027909	-77.119206	59.63
2	<0.01	1.88	31.24	42.901427	-79.618196	55.92
3	<0.01	2.24	28.93	44.596204	-80.794748	36.41
4	<0.01	3.29	19.33	45.100714	-81.277772	19.41
5	0.04	4.39	14.09	45.129385	-76.124550	3.56
6*	0.051	2.20	13.48	42.880048	-75.031825	21.29

\*Secondary cluster

Allevi et al., 2013; Knobloch et al., 2013; Swistock et al., 2013). This is the first major bacteriological spatial analysis of private well water submissions. The maximum sample size of previous bacteriological well quality studies was limited to 4,000 wells from the state of Wisconsin (Knobloch et al., 2013). Hunter et al. (2011) studied 14,000 private water supplies (servicing less than 50 people) in England and France (2011). Additionally, a study by Perkins et al. (2009) investigated the quality of rural tap water of 5,000 dairy farms in all of southern Ontario using spatial analysis techniques. Conversely, the magnitude of this study is far greater than that of previous work with a total of 90,000 samples representing 30,000 private wells in a subsection of southern Ontario. By combining spatial analyses with a bacteriological quality investigation, and utilizing a greater power of detection, areas of high contamination could be identified.

This investigation revealed several clusters with a significantly higher risk of *E. coli* contamination for private well water submissions at the regional level in southeastern Ontario. One sizeable (at the 50% level) and four smaller distinct (at the 5% level) clusters were observed over the 5-year period investigated (2008-2012). Considering that private well water submission locations are not identical per annum, this finding implies that these contamination clusters are spatially stable. Furthermore, this suggests that there are underlying factors that contribute to the higher risk of *E. coli* contamination in these identified cluster regions.

The 5-year, geographically stable *E. coli* cluster in HPEC identified by this study was also discovered in a study by Perkins et al. (2009), which investigated *E. coli* contamination of rural milk house tap water for the 2003-2004 period. Both Ontario farms and private wells are located in rural areas supplied by groundwater; therefore regional hydrogeology may be a contributing factor to the contamination. The HPEC contamination cluster was located in a limestone region with limited overburden, and previous studies

have noted that geologically similar areas have a high susceptibility to contaminants (Levison and Novakowski, 2008; Hynds et al., 2012; Swistock et al., 2013). Other temporally stable factors have been described as affecting rural water supplies. That are also temporally stable, such as socioeconomic status (US Environmental Protection Agency, 1984; Evans and Kantrowitz, 2002). Previous studies have also investigated the relationship among education, private well water stewardship and poor outcomes including lack of well disinfection, testing and maintenance (Kreutzwiser et al., 2010, 2011). Additionally, physical well attributes could be a factor as previous studies show that poorly constructed, shallower and older wells are most susceptible to contamination (Goss et al., 1998; Kreutzwiser et al., 2010; Hynds et al., 2012; Allevi et al., 2013; Swistock et al., 2013). Land use has also been implicated, especially with chemical and microbiological contaminants from agricultural runoff (Goss et al., 1998; Rudolph et al., 1998; Benson et al., 2006; Allevi et al., 2013). Consequently, any of these factors may contribute to the clusters observed in this study.

Four distinct smaller cluster regions were also detected, which showed temporal shift and which were not represented every year. This movement and variation of contamination may be attributed to the presence of variable factors, for example precipitation, whose effect would be accentuated by the underlying geology. Since weather events tend to be localised, the cluster patterns observed may be related to the precipitation severity and location for a given year. Other studies have shown that heavy rainfall, when combined with porous underlying geology have adverse groundwater effects (Arnade, 1999; Curriero et al., 2001; Bacci and Chapman, 2011; Hynds et al., 2012; Swistock et al., 2013). All of the regions identified can be targeted for public health initiatives and interventions to reduce the potential human health risk consequent to drinking well water contamination.

This study found three other major clusters (one located in Niagara and two in the Bruce Peninsula) for the southern Ontario region in 2012 in addition to the regional HPEC cluster. A fourth area of two clusters also existed near the Ottawa area; however, these clusters were significantly smaller and had P-values several orders of magnitude higher than the rest. A limitation of the provincial analysis was the use of a single year of data. However, our results corroborate a previous spatial analysis study performed on tap water from cattle milk houses in southern Ontario (Perkins et al., 2009) where the same three primary cluster regions (Niagara, HPEC and Bruce Peninsula) were observed. The milk house study was conducted between 2003 and 2004 and used a different sample type (Perkins, 2006), which further strengthens these findings. A map comparing the clusters detected by this study and the milk house study can be found in Fig. 6. The provincial results further implicate stable contributing factors to *E. coli* contamination of rural water supplies. Defining *E. coli* contamination patterns at a finer scale for the rest of southern Ontario could better inform public health and policy agencies enabling direct interventions such as proper well stewardship, education and surveying.

The water submissions used by this study were a convenience sample and as such some of the samples may have not originated from private well sources but rather from municipal sources or surface water. Additionally, some of the samples may be from well

systems with private treatment systems. However, given the magnitude of the sample size, this should not have a significant impact on the results found and both municipal and treated well water would favour the null hypothesis (Uhlmann et al., 2009).

A spatial scan statistic was appropriately applied to this study given that it is able to utilise point-level data. Other area-based measures of spatial association, such as local indicators of spatial association (LISA), require aggregating the point-level data, which would incur a loss of spatial precision and therefore have less power to detect clusters of contaminated wells. Furthermore, as demonstrated by Song and Kulldorff (2003), the spatial scan statistic has detection power comparable to Besag-Newell's R and Cuzick-Edwards' k-NN tests, but the power of those tests is highly dependent on the choice of parameter values and perform well when the size or scale of clustering is known. In this exploratory study, we had no *a priori* hypotheses about the scale of spatial clustering, and therefore these tests would have been inappropriate. A limitation of the spatial analysis software utilised was the circular scanning window as it is possible that clusters exist, which cannot be detected using a circular contour. However the use of a range of sizes of scanning windows minimises this risk (Chen et al., 2008). Furthermore, while non-circular algorithms have been developed they were not feasible to apply in this case due to the magnitude of data. For example, Tango and Takahashi (2005) developed a flexibly shaped scan statistic, but noted that it had lim-



Fig. 6. Comparison between provincial 2012 data and Perkins' (2006) Southern Ontario data on *E. coli* contamination in milk house water from 2003-2004. Orange - our study 2012 data; Grey - Perkins 2003-2004 data.

ited power of detection, was unable to identify secondary clusters, and was only appropriate for small to moderate cluster sizes. Therefore a circular scanning window was deemed sufficient as the objective for this study was to identify “hotspots” of contamination. Further studies will be undertaken to more definitively delineate the boundaries of identified patterns of contamination now that the circular scan statistic has revealed the major regions of concern.

One of the many strengths of the study was the high match rate achieved by the unique geocoding process. Rural large-scale investigations at the address level are infrequent given the labour intensive nature of the geocoding processes involved. Rural addresses are complicated and multifaceted especially in North America where major changes to address systems are underway (Goldberg, 2008). This study successfully geocoded 90,149 private well water addresses to 30,687 unique locations out of 107,547 private well water submissions for the 2008-2012 time period obtaining a match rate of 84%.

The consumption of compromised water is a global threat to human health (Furtado et al., 1998; Raina et al., 1999; Macler and Merkle, 2000; Corkal et al., 2004). This paper demonstrates that 5.2% of wells tested in Southern Ontario in 2012 had at least one *E. coli* - positive result. Protecting rural populations (30% of Canada) from water-borne illness requires a prior knowledge of at risk regions. The findings of this study substantially extend the current private well water research both nationally and globally by demonstrating the presence of *E. coli* contamination hotspots. Furthermore, the methodologies used in this study can be applied to any region where private well water sources are prevalent. For example, in the United States alone, private wells supply drinking water to 13 million households (United States Census Bureau, 2010) and in Ireland, 17% of the population rely on private groundwater sources (Hynds et al., 2012).

Generally, public health spatial analysis studies have utilised single time frames (Odoi et al., 2004; Green et al., 2006). A study by Pearl et al. (2006), noted the importance of differentiating between time periods to track any contamination patterns and elucidate underlying causes. This study spatially analysed the regional dataset individually by year and compared results across the 5-year time frame. Therefore we identified both spatially stable and variable clusters that assist with future proactive investigations into possible contributing factors to faecal contamination of private well water sources with the hope that preventive strategies can be discovered and applied.

## Conclusion

We have demonstrated the existence and extent of a persistent drinking water quality issue in Southern and Southeastern Ontario that suggests the presence of underlying factors. The spatial analytics applied here cannot reveal the specific primary and contributing mechanisms of *E. coli* contamination, but can provide the groundwork for future investigation.

## Acknowledgements

The authors wish to thank the employees of the Public Health Ontario Laboratories in Kingston for their kind support. Special thanks go to Kristi Randhawa, Lydia Cheng, Susan Greaves, Kayla Poirier and Fran Lao.

## References

- Allevi RP, Krometis LAH, Hagedorn C, Benham B, Lawrence AH, Ling EJ, Ziegler PE, 2013. Quantitative analysis of microbial contamination in private drinking water supply systems. *J Water Health* 11, 244-255.
- Arnade LJ, 1999. Seasonal correlation of well contamination and septic tank distance. *Ground Water* 37, 920-923.
- Bacci F, Chapman DV, 2011. Microbiological assessment of private drinking water supplies Co. Cork, Ireland. *J Water Health*, 738-751.
- Benson VS, VanLeeuwen JA, Sanchez J, Dohoo IR, Simers GH, 2006. Spatial analysis of land use impact on ground water nitrate concentrations. *J Environ Qual* 35, 421-432.
- Brownstein JS, Rosen H, Purdy D, Miller JR, Merlino M, Mostashari F, Fish D, 2002. Spatial analysis of West Nile virus: rapid risk assessment of an introduced vector-borne zoonosis. *Vector-Borne Zoonot* 2, 157-164.
- Charrois JWA, 2010. Private drinking water supplies: challenges for public health. *CMAJ* 182, 1061-1064.
- Chen J, Roth RE, Naito AT, Lengerich EJ, Maceachren AM, 2008. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *Int J Health Geogr* 7, 57-74.
- Coleman M, Mabuza AM, Kok G, Coetzee M, Durrheim DN, 2009. Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. *Malar J* 8, 68.
- Corkal D, Schutzman WC, Hilliard CR, 2004. Rural water safety from the source to the on-farm tap. *J Toxicol Env Heal A* 67, 1619-1642.
- Curriero FC, Patz JA, Rose JB, Subhash L, 2001. The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994. *Am J Public Health* 91, 1194-1199.
- Evans GW, Kantrowitz E, 2002. Socioeconomic status and

- health: the potential role of environmental risk exposure. *Annu Rev Publ Health* 23, 303-331.
- Furtado C, Adak GK, Stuart JM, Wall PG, Evans HS, Casemore D, 1998. Outbreaks of waterborne infectious intestinal disease in England and Wales, 1992-5. *Epidemiol Infect* 121, 109-119.
- Goldberg DW, 2008. A geocoding best practices guide. North American Association of Central Cancer Registries, 287 pp.
- Goss MJ, Barry DAJ, Rudolph DL, 1998. Contamination in Ontario farmstead domestic wells and its association with agriculture: 1. Results from drinking water wells. *J Contam Hydrol* 32, 267-293.
- Green CG, Krause DO, Wylie JL, 2006. Spatial analysis of *Campylobacter* infection in the Canadian province of Manitoba. *Int J Health Geogr* 5, 2.
- Hunter PR, de Saylor MA, Risebro HL, Nichols GL, Kay D, Hartemann P, 2011. Quantitative microbial risk assessment of Cryptosporidiosis and Giardiasis from very small private water supplies. *Risk Anal* 31, 228-236.
- Hynds PD, Misstear BD, and Gill LW, 2012. Development of a microbial contamination susceptibility model for private domestic groundwater sources. *Water Resour Res* 48, 1-13.
- Knobelech L, Gorski P, Christenson MS, Anderson H, 2013. Private drinking water quality in rural Wisconsin. *J Environ Health* 75, 16-20.
- Kreutzwiser R, de Loë RC, Imgrund K, 2010. Out of sight, out of mind: private water well stewardship in Ontario. Report on the findings of the Ontario household water well owner survey 2008. Waterloo, Ontario. University of Waterloo, 164 pp.
- Kreutzwiser R, de Loe R, Imgrund K, Conboy MJ, Simpson H, Plummer R, 2011. Understanding stewardship behaviour: factors facilitating and constraining private water well stewardship. *J Environ Manage* 92, 1104-1114.
- Kulldorff M, 1997. A spatial scan statistic. *Commun Stat Theory* 26, 1481-1496.
- Levison J, Novakowski K, 2009. The impact of cattle pasturing on groundwater quality in bedrock aquifers having minimal overburden. *Hydrogeol J* 17, 559-569.
- Macler BA, Merkle JC, 2000. Current knowledge on groundwater microbial pathogens and their control. *Hydrogeol J* 8, 29-40.
- Odoi A, Martin SW, Michel P, Middleton D, Holt J, Wilson J, 2004. Investigation of clusters of giardiasis using GIS and a spatial scan statistic. *Int J Health Geogr* 3, 11-21.
- Ontario Ministry of the Environment, 1994. Ontario's drinking water objectives. Toronto, Ontario. Queen's Printer for Ontario.
- Ontario Ministry of the Environment, 2009. Clean water act. Toronto, Ontario. Queen's Printer for Ontario.
- Pearl DL, Louie M, Chui L, Dore K, Grimsrud KM, Leedell D, Martin SW, Michel P, Svenson LW, McEwan SA, 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.
- Perkins NR, 2006. The impact of wash water quality and milk temperature monitoring on milk quality on Ontario dairy farms. The University of Guelph, Guelph Ontario UMI MR25467.
- Perkins NR, Kelton DF, Hand KJ, MacNaughton G, Berke O, Leslie KE, 2009. An analysis of the relationship between bulk tank milk quality and wash water quality on dairy farms in Ontario, Canada. *J Dairy Sci* 92, 3714-3722.
- Raina PS, Pollari FL, Teare GF, Goss MJ, Barry DA, Wilson JB, 1999. The relationship between *E. coli* indicator bacteria in well-water and gastrointestinal illnesses in rural families. *Can J Public Health* 90, 172-175.
- Rudolph DL, Barry DAJ, Goss MJ, 1998. Contamination in Ontario farmstead domestic wells and its association with agriculture: 2. Results from multilevel monitoring well installations. *J Contam Hydrol* 32, 295-311.
- Said B, Wright F, Nichols GL, Reacher M, Rutter M, 2003. Outbreaks of infectious disease associated with private drinking water supplies in England and Wales 1970-2000. *Epidemiol Infect* 130, 469-479.
- Salvadori MI, Sontrop JM, Garg AX, Moist LM, Suri RS, Clark WF, 2009. Factors that led to the Walkerton tragedy. *Kidney Int* 75, S33-S34.
- Schuster CJ, Ellis AG, Robertson WJ, Charron DF, Aramini JJ, Marshall, BJ, Medeiros DT, 2005. Infectious disease outbreaks related to drinking water in Canada, 1974-2001. *Can J Public Health* 96, 254-258.
- Sierra Legal Defense Fund, 2006. Waterproof 2: Canada's drinking water report card. Available at: <http://www.ecojustice.ca/publications/reports/waterproof-2/attachment> (accessed on June 2013).
- Song C, Kulldorff M, 2003. Power evaluation of disease clustering tests. *Int J Health Geogr* 2, 9.
- Smith A, Reacher M, Smerdon W, Adak GK, Nichols G, Chalmers RM, 2006. Outbreaks of infectious intestinal disease in England and Wales, 1992-2003. *Epidemiol Infect* 134, 1141-1149.
- Statistics Canada, 2011. Households and the environment. Ottawa; 2011. Available at: <http://www.statcan.gc.ca/pub/11-526-x/11-526-x2011001-eng.htm> (accessed on June 2013).
- Strauss B, King W, Ley A, Hoey JR, 2001. A prospective study of rural drinking water quality and acute gastrointestinal illness. *BMC Public Health* 1, 8.
- Swistock BR, Clemens S, Rummel S, 2013. Water quality and management of private drinking water wells in Pennsylvania, 2013. *J Environ Health* 75, 60-66.
- Szonyi B, Wade SE, Mohammed HO, 2010. Temporal and spatial dynamics of *Cryptosporidium parvum* infection on dairy farms in the New York City watershed: a cluster analysis based on crude and Bayesian risk estimates. *Int J Health Geogr*

- 9, 31.
- Tango T, Takahashi K, 2005. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr* 18, 4.
- Uhlmann S, Galanis E, Takaro T, Mak S, Gustafson L, Embree G, Bellack N, Corbett K, Isaac-Renton J, 2009. Where's the pump? Associating sporadic enteric disease with drinking water using a geographic information system, in British Columbia, Canada, 1996-2005. *J Water Health* 7, 692-698.
- United States Census Bureau, 2010. American housing survey for the United States: 2009-current housing report. Available at: <http://www.census.gov/prod/2011pubs/h150-09.pdf> (accessed on June 2013).
- US Environmental Protection Agency, 1984. National Statistical Assessment of Rural Water Conditions. Washington, DC.
- WHO, 2010. WHO guidelines for drinking-water quality. Geneva: World Health Organization.