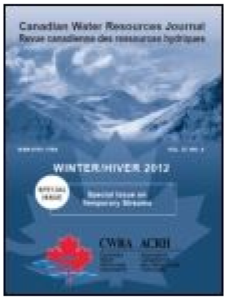


Escherichia coli contamination of rural well water in Alberta, Canada is associated with soil properties, density of livestock and precipitation

Journal:	*Canadian Water Resources Journal*
Manuscript ID	TCWR-2018-0062.R2
Manuscript Type:	Original Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Invik, Jesse; University of Calgary, Ecosystem and Public Health Barkema, Herman; University of Calgary, Department of Animal Production Health, Faculty of Veterinary Medicine; University of Calgary, Cumming School of Medicine Massolo, Alessandro; University of Calgary, Department of Ecosystem and Public Health, Faculty of Veterinary Medicine; Universita di Pisa, Ethology Unit Neumann, Norman; University of Alberta, School of Public Health Cey, Edwin; University of Calgary, Department of Geoscience Checkley, Sylvia; University of Calgary, Ecosystem and Public Health, Faculty of Veterinary Medicine; University of Calgary, Cumming School of Medicine; Alberta Provincial Laboratory for Public Health, Calgary Laboratory Site
Keywords:	<i>Escherichia coli</i>, groundwater, geographic information systems, generalized estimating equations, Alberta

Title: *Escherichia coli* contamination of rural well water in Alberta, Canada is associated with soil properties, density of livestock and precipitation

Jesse Invik [a], Herman W. Barkema [b,c], Alessandro Massolo [a,c*], Norman F. Neumann [d],
Edwin Cey [e], Sylvia Checkley [a,c,f**]

[a] *Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, 3280 Hospital Dr. NW, Calgary, AB, T2N 4Z6, Canada*

[b] *Department of Production Animal Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, 3330 Hospital Dr. NW, AB, T2N 4N1, Canada*

[c] *Cumming School of Medicine, University of Calgary, 3280 Hospital Dr. NW, Calgary, AB, Canada T2N 4Z6, Canada*

[d] *School of Public Health, University of Alberta, 3-300 Edmonton Clinic Health Authority, 11405-87 Ave Edmonton, AB, T6G 1C9, Canada*

[e] *Department of Geoscience, University of Calgary, 2500 University Dr. NW, Calgary, AB, T2N 4N1, Canada*

[f] *Alberta Provincial Laboratory for Public Health, Calgary Laboratory Site, 3030 Hospital Dr. NW, Calgary, AB, T2N 4W4, Canada*

**A. Massolo is now at Ethology Unit, Department of Biology, Università di Pisa, Pisa, Italy.*

***Corresponding author, Sylvia Checkley, Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, 3280 Hospital Dr. NW, Calgary, AB, T2N 4Z6, Canada, slcheckl@ucalgary.ca*

ABSTRACT

Waterborne outbreaks of infectious disease continue to be a public health risk, particularly those in areas where testing of private and public small system groundwater systems is left to the owners/overseers of these wells who may not recognize the importance of testing and treatment. Recognizing factors associated with contamination of wells is important for public safety and can encourage well owners/overseers to test regularly and properly maintain drinking water supplies. Tests results for presence/absence of total coliforms and *Escherichia coli* for private and public untreated well water for the years 2010-2012 (n=56,609) were provided by the Alberta Provincial Laboratory for Public Health. Tests were geolocated with the Alberta Township Survey System and aggregated to the quarter section. Agricultural independent variables were provided by the Canadian Agricultural Census and monthly cumulative precipitation was calculated using Alberta Agriculture and Forestry's website of weather station data. Overall frequency of *E. coli*-positive wells in the study was 1.4%. A marginal multivariable logistic regression model was fit using generalized estimating equations to account for repeat testing of some quarter sections. Three significant factors associated with increased *E. coli*-positive untreated drinking water wells were identified: soil properties (KSat and sand content), animal density and monthly cumulative precipitation.

RÉSUMÉ

Les maladies infectieuses reliée à l'eau continuent de présenter un risque a la santé publique. Ceci particulièrement dans les endroits ou les tests de la qualité de l'eau sont laissés aux mains des propriétaires des puits artésiens, qui peuvent parfois être dans l'ignorance dans le domaine des tests et du traitement de l'eau. Reconnaître les facteurs associés à la contamination pour les puits est une facette importante pour la santé publique et peux emmener les propriétaires de ceux-ci a procéder a des test réguliers pour aider à maintenir l'approvisionnement en eau potable. Les résultats concernant la présence ou l'absence de coliformes et de l'*Escherichia coli* pour les puits privé et les puits d'eau non-traitée publique pour les années 2010-2012 (56,609) furent obtenu du laboratoire pour la santé publique de la province de l'Alberta. Les tests utilisaient des données de géolocation des contés de l'Alberta et organisé en quarts de sections. Les indépendant agricoles furent obtenus au sein du recensement agricole du Canada et les précipitations cumulatives mensuelles furent obtenu depuis le site internet des stations météorologiques du ministère de l'agriculture et de la foresterie de l'Alberta. Un modèle de régression logistique multivariable marginal fut installé en utilisant des équations d'estimation générales pour prendre en considération les tests répétitifs de certaines propriétés. Au total, la fréquence de tests positifs pour *l'E Coli* dans cette étude est de 1.4%. Trois facteurs importants associés à l'augmentation du nombre de puits d'eau potable non traités pour E. loci ont été identifiés: propriétés du sol (de conductivité hydraulique et teneur en sable), les animaux en pâturage, et l'accumulation des précipitations mensuelles.

Keywords: *Escherichia coli*; groundwater; geographic information systems; generalized estimating equations; Alberta; drinking water

Introduction

Worldwide, over 1.8 billion people access drinking water contaminated with faecal material. Unsafe water causes over 500,000 deaths due to diarrhea annually (WHO 2016). Although considered safer than surface water, groundwater contamination of drinking water with pathogenic bacteria is a public health issue worldwide (Garvey et al. 2016; Liu XiaoQing et al. 2015). North America is not exempt from this public health issue, and disease outbreaks from untreated drinking water are not uncommon in Canada and the United States (Hynds et al. 2014; Wallender et al. 2014). A 2016 study estimated that over 100,000 cases of acute gastrointestinal illness were due to pathogens in small private water systems and small communal systems in Canada, which would be primarily ground water systems (Murphy et al. 2016).

Drinking water for over 400,000 Albertans is supplied by private wells or small communal systems (Government of Alberta n.d.; Summers 2010). Much of the water consumed from private wells and small communal systems is untreated, and maintenance and monitoring falls to the discretion of the well owner/overseer. A 2010 survey of well owners in Alberta reported that only 11% of private well owners tested their wells at least annually for bacterial quality indicators and 70% felt maintenance and decontamination were unnecessary except in response to a problem. Some of the reasons people may not test and/or treat well water included a lack of understanding about ground water. For instance more than 50% of respondents in the study did not believe that a deeper well offered safer water. A high level of confidence in the safety of wells was coupled with a lack of any kind of testing of those same wells. A majority of respondents felt that it was better not to perform preventative maintenance on their wells unless motivated by a specific reason, such as health concerns or concerns over events such as flooding (Summers 2010). A similar study from Wisconsin identified that only 10% of respondents had tested their well in the previous year, and reasons for not testing, not performing maintenance were also similar (Malecki et al. 2017). As previous outbreaks have demonstrated, waterborne pathogens are a risk to public health and incidence of waterborne disease may even be increasing (Neumann et al. 2005). In addition, climate change is predicted to increase evaporation and precipitation with increasing storms, storm intensity and flooding in some regions (Dettinger 2011; Scaife et al. 2012). Understanding what factors are associated with contamination can help motivate well owners to make smart decisions concerning the maintenance and monitoring of their water supply.

Coliforms have been used as an indicator for microbial water quality since the 1880's (Gleeson and Gray 1996). *Escherichia coli* is recommended as an indicator of fecal contamination by the World Health Organization (WHO), however with caveats and recommendations to include other indicators depending on the particular conditions and threats (World Health Organization 2011). As laboratory techniques such as polymerase chain reaction (PCR) become more commonly used and affordable, direct pathogen testing may become more common. In the meantime, the use of *E. coli* as an indicator remains valuable.

A number of factors associated with total coliforms and *E. coli* contamination of rural well water have been identified. Cattle and other ruminants are well documented as reservoirs for pathogenic and non-pathogenic *E. coli* (Arthur et al. 2009; Berry et al. 2010; Karmali 2002), and there has been an association between areas where animal husbandry is combined with manure spreading, and higher incidence of pathogenic human *E. coli* cases (Michel et al. 1999; Valcour et al. 2002). Other livestock can also harbour pathogenic as well as non-pathogenic *E. coli*, including pigs, sheep, and chickens (Bergeron et al. 2012; Dassouli-Mrani-Belkebir et al. 1988; Deckert et al. 2010). A seasonal trend has been reported both in human cases of *E. coli* O157:H7 (peak in July) (Michel et al. 1999), and in prevalence of *E. coli* O157 in rectal samples of cattle (peaks in Spring and late Summer) (Chapman et al. 1997). According to the 2011 census of agriculture in Canada, Alberta has 40% of the nation’s cattle (5,104,605 head), as well as 37% of llamas and alpacas (11,740), 46% of domestic bison (57,483) and 52% of domestic elk (16,286)(Government of Canada 2012). These animals represent potential sources of *E. coli* and TC contamination of water. In addition, zero-tillage farming,the practice of not breaking up the soil yearly with a tiller to better maintain moisture content, (Mitchell et al. 2012), can lead to the creation of macropores which allow water to by-pass the filtering action of soil and may lead to contamination of groundwater supplies (Samarajeewa et al. 2012; Thiagarajan et al. 2007; Krog et al. 2017). Extreme weather events have also been linked to waterborne pathogen outbreaks, as rapid transport of pathogens can be increased when soils become heavily saturated, potentially transporting pathogens into water sources (Cann et al. 2013;

Curriero et al. 2001). Human sources of fecal material also may impact water quality and human health (Denno et al. 2009).

The objectives of the study were to determine if there is an association between agricultural practices, soil types and precipitation patterns with the contamination of rural well water with *E. coli* and TC.

Methods

Study Region

The study region for this project focuses on the Central region of the Province of Alberta, Canada (Figure 1) to enhance generalizability of results. The Central region has the highest numbers of water quality tests in the province performed by the Alberta Provincial Laboratory for Public Health (ProvLab) and has areas of both high and low intensity of agriculture. The counties containing National Parks (Banff and Jasper) were excluded because water quality tests from Federal Parks were not available for this study. Counties including Calgary and Edmonton, the two largest cities in the province and the town of Drumheller, were also excluded as these counties represent (by percentage of land mass) primarily urban areas whereas other included counties may include cities or towns but are primarily rural (by percentage land mass) and these urban areas would decrease generalizability of study results.

Figure 1: Study area (shaded), with counties demarcated with black outlines for Alberta, Canada.

Samples

Water submission data were provided by ProvLab for the years 2010-2012 and were accessed using the Data Integration for Alberta Laboratories (DIAL) tool, a web-based surveillance tool developed by ProvLab (Mukhi et al. 2011). The data included 56,609 tests of rural groundwater samples, both public and private. Public/communal systems included non-transient systems (<15 connections) and transient systems such as community halls and campgrounds as defined by Alberta Environment (Alberta Environment and Parks 2009). All water samples were provided voluntarily by the well owner/overseer. Samples were excluded from the study if they were missing the Alberta Township Survey System (ATS) location or if the attributes of this system were incorrect making geolocation impossible, and if the location of the well was outside of the study area.

The Alberta Township Survey System was used to assign geographical coordinates to the submission data. This information allows a parcel of land to be located to a resolution of one quarter section ($0.65km^2$) (Alberta Environment and Parks 2010). This information was voluntarily provided by a portion of the water sample submitters (70%) at the ProvLab. In some cases not all attributes of the ATS system were provided or one or more attributes were incorrect. Samples with missing or invalid ATS attributes were excluded from further analysis. Approximately 66% of samples could be geolocated using the ATS system and aggregated to the level of the quarter section, and testing of these samples for TC and *E. coli* contamination provided the two outcome variables for modelling. In some cases the quarter section could contain multiple wells, but each quarter section was treated as though all tests from a single quarter section were from the same well. Multiple tests from the same quarter section were treated as clusters in the model.

Laboratory Analyses

Microbiological water quality testing for *E. coli* and TC was performed by ProvLab, an ISO 17025 accredited laboratory. As part of routine provincial water testing, water samples were collected in 250 mL sterilized plastic vessels and delivered to ProvLab's Environmental Waters Laboratory within 24 hours of collection where 100 mL were tested for TC and *E. coli* using an enzyme substrate test (Colilert®, IDEXX Laboratories, Westbrook, ME, USA) according to the manufacturer's protocol (IDEXX Laboratories 2012). In the case of a positive sample, local public health agencies informed the well owner/overseer of the microbiologically contaminated drinking water samples and provide further information regarding re-testing/decontamination as per Alberta Health and Wellness's *Environmental Public Health Field Manual* (Technical Advisory Committee on Safe Drinking Water 2007).

Independent Variables

Four groups of independent variables were identified for inclusion in the study; agricultural, temporal, precipitation, and soil properties. Agricultural variable information was extracted from the Canadian Census of Agriculture for 2011 and aggregated to census regions (Government of Canada 2012). There were 39 agricultural census regions included in the study area, with a median size of 4,278 km^2 (ranging from 1,308 to 19,038 km^2). Four agricultural variables were selected for possible inclusion into the models: 1) density of large animals (cattle, sheep, horses, goats, lamas, bison, elk, and deer) excluding pigs (animals/km^2), 2) percentage of land devoted to zero-till farming, 3) percentage of land with manure applied, and 4) percentage of land devoted to animal grazing. The Geospatial Modelling Environment (GME) software (version 0.7.2.1; Beyer, 2012) was used to extract values for the centroid of each quarter section. A complete list of variable categories is provided with the full model output in Appendix A.

Year and month of collection of the sample was included in the models to account for different temperature/precipitation patterns over time. Precipitation data were collected from Alberta Agricultural and Forestry using daily precipitation data from all Alberta weather stations (Alberta Agriculture and Forestry 2014). Only data from weather stations within and close to the study area (within 100 km of the boundaries of the study area) and those that had complete data during the study time period were considered. Monthly precipitation accumulation (mm) was calculated for each weather station and data extrapolation to the whole study area was performed to obtain continuous data for each quarter section using an ordinary kriging interpolation method implemented in ArcGIS® (version 10.1, ESRI 2012, Redlands, CA, USA). Outliers were removed before kriging was performed for each layer. Using this method, one raster layer was created for each month of the study, showing accumulated precipitation for that month (Cell resolution 2.063 km). The GME software (version 0.7.2.1, Beyer, 2012) was then used to extract the accumulated precipitation monthly value from the temporally appropriate raster layer for the geographical location of each water sample. When the water sample was collected in the first half of the month, the previous month’s precipitation layer was used for extraction, and when the water sample was collected in the second half of the month, the same month’s precipitation layer was used for extraction.

Soil data were provided by Agriculture and Agri-Food Canada through the Soil Landscapes of Canada database, version 3.2, 2011 (Agricultural and Agri-Food Canada 2008). Three soil variables were selected for possible inclusion into the models: soil saturated hydraulic conductivity (KSat) in cm/hour, average soil clay content in percent and average soil sand content in percent. The GIS layers included with the database were based on previous soil survey maps at a scale of 1:1 million (Agricultural and Agri-Food Canada 2008). Attribute tables for each polygon included a hierarchical association between the polygon and one or more soil components with each component associated with one or more soil layers. Clay content, sand content and KSat were provided at the level of soil layers. For each soil component, the average vertical hydraulic conductivity across all soil layers was calculated as follows:

$$K = \frac{b}{\sum_{i=1}^{n} b_i / K_i}$$

K=average hydraulic conductivity (cm/hour)
K_1, K_2 . . . K_n hydraulic conductivity of individual layers
b_i = thickness of layer i
b=total thickness of all layers

After calculation for each soil component, an areal average for all soil components within the polygon was calculated as follows:

$$K = (K_1^{A_1} \times K_2^{A_2} \times \ldots K_n^{A_{1n}})^{1/A}$$

K = average vertical hydraulic conductivity
K_1, K_2,..K_n = hydraulic conductivity of individual components 1, 2,…,n
A_i = percentage for soil component ‘i’
A = total percentage of all soil regions = A_1 +A_2+ …+A_n

Calculation of average sand and clay content for each polygon was calculated as follows:

$$A_c = \frac{\sum_{i=1}^{n}(b_i \times A_i)}{b}$$

A_c = average percent clay/sand for soil component
A_1..A_n = average sand/clay for individual layers 1, 2,…,n
b_i = thickness of layer i
b=total thickness of all layers

Following this calculation, the simple average of the clay/sand for the different soil components in the polygon was calculated.

Statistical Analysis

Data analysis was conducted using SPSS (version 22.0.0, IBM Corp© 2013), and statistical significance was set at p<.05. The relationship between all independent variables and the logit of the outcome variables was not linear (with the exception of precipitation for *E. coli* contamination) by visual inspection. Data transformations of the independent variables were attempted including log-normal (log10(x)), square root ($\sqrt{x}$) and inverse transformation (1/x). Data transformations did not improve the relationship between the logit of the outcome variables and the independent variables, so variables were categorized using quartiles to create categories in terms of number of samples in each category, with data ties (cases with the same value for the independent variable) going into the higher category. Correlation between the independent variables was tested in the categorical form using Cramer's V statistic, and values of > 0.6 were considered unacceptably high, leading to removal of one or more variables from the model.

Total coliform and *E. coli* positive samples were clustered by location, so logistic regression using generalized estimating equations (GEE) (Zeger and Liang 1986) was used for modelling. The first order autoregressive (AR-1) correlation structure was used because wells that were tested multiple times were more likely to have dependent results the closer in time the tests were to each other. The AR-1 correlation structure appropriateness was also evaluated against exchangeable and unstructured correlation structures using the quasi-likelihood criteria in the intercept-only models as well as the final models (Pan 2001). The Hosmer and Lemeshow seven-step procedure for logistic regression model building (modified for GEE) was used for variable inclusion in the model using the quasi-likelihood criteria for variable selection (Hosmer et al. 2013; Pan 2001). Once the main model was decided on, interactions were introduced one at a time from a list of pairs that had a reasonable and biologically informed chance of interaction and these pairs must have been significant at p=.05 to remain in the model. Before the model was accepted, model fitness was checked (Hosmer et al. 2013). The equation for predicting the probability of Y=1 or a positive test for *E. coli* or total coliforms takes the following form (Horton et al. 1999), where x_{ij} refers to the value of x for sample i in cluster j.

$$P(Y_{ij} = 1|x_{ij}) = \frac{\exp\,(\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_k x_{kij})}{1 + \exp\,(\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_k x_{kij})}$$

Year and month of collection of the sample were included in the models to account for different temperature/precipitation patterns over time. For temporal variables, month and year, the month or year with the lowest marginal mean for *E. coli*/TC contamination was chosen as the reference month. Model goodness of fit was evaluated by creating groups based on the deciles of estimated probability for the outcome variable, and then expected and observed outcome were evaluated in a contingency table with the Hosmer and Lemeshow test statistic for Goodness of Fit to test if there was any significant lack of fit between the data and the model output (Hosmer et al. 2013) (Evans and Li 2005). In addition, predictive efficacy of the models was examined using the receiver sensitivity operating characteristic curve, area under the curve statistic (ROC-AUC) (Hosmer et al. 2013). The ROC curve was calculated using the predicted value of the mean of the response as the test variable and the outcome variable (TC or *E. coli* positivity) as the state variable. Results were expressed in the form of an odds ratio (OR) which is a measure of association between a factor and the outcome, when controlling for other factors in the model (Szumilas 2010).

Results

Outcome variable

A total of 56,609 test results for TC and *E. coli* were provided by ProvLab for the years 2010-2012. Of these tests, 19,481 were excluded because ATS coordinates were missing or incorrect and it was not possible to geolocate the tests. A further 12,708 were outside the study area. The study time period was then further restricted to water submissions made during the months when the ground in Alberta is typically not frozen (April-October) to simplify issues around precipitation. Precipitation falling as snow and accumulating on the ground rather than flowing down through soil would contribute differently to contamination of groundwater than rainfall. This left 17,237 water submission results, distributed over 7,728 unique quarter sections in the study. Multiple tests on wells on the same quarter section of land resulted in clustered data. These tests may or may not have been from the same well. The minimum cluster size was one and maximum size was 326, with a median of 1 and an interquartile range of 1. Ignoring the clusters of submissions within quarter sections, 1.7% of all tests were positive for *E. coli* and 18.6% of all tests were TC-positive (16.9% were TC-positive but negative for *E. coli*).

Independent variables

Agricultural variables, % land devoted to grazing and % of land with manure applied, were both positively associated with density of large animals (Cramer's V =.60, P<.001 and Cramer's V =.72, P<.001, respectively). All three variables are indicators of the amount of fecal material on the land. Given the value of Cramer's V, density of large animals was arbitrarily chosen as a proxy of all three. A similar decision was made for the soil variables. Average clay content was inversely associated with average sand content (Cramer's V = 0.60, P<.001), but both were less highly associated with KSat (clay, Cramer's V = 0.44, P<.001, sand, Cramer's V = 0.45, P<.001). Average clay soil content was discarded and average sand content and KSat were retained.

Total coliform models

In the TC model, all independent variables, except for KSat and sand were significant in the unconditional logistic regression analyses (Table A.1) with a significance of .05. All variables were associated with TC positivity at p<.25, and so were therefore included in the initial multivariate model building process (Hosmer et al. 2013).

The final multivariable TC model (Table A.2) included the following variables: precipitation, zero-till farming, sand, density of large animals, KSat, year and month. No significant interaction terms were identified for inclusion (Table A.3). Significant associations in the model included precipitation overall, though no individual categories were significant. The odds of TC contamination was 1.29 times (95% CI 1.05-1.59, p=.02) higher for zero till farming was between 24.1 and 61.9% compared to the reference category, 0-5.1%, adjusting for all other factors in the model. Sand was associated with TC-positivity in the model overall (Wald Chi-Square=11.006, df=3, p=.01), and all categories were significant compared to the reference categories, with the odds of contamination between 1.22 and 1.38 times higher compared to the reference level, when adjusting for all other factors in the model. Only one category of density of large animals was significant, with large animal densities of 26.3 to 32.7 animals/km^2 actually associated with decreased TC contamination in comparison to the reference level of less than 14.9 animals/km^2 when adjusting for all other factors in the model (OR=0.77, 95% CI 0.63-0.93, p=.008). The odds of TC contamination was 1.31 to 1.48 times higher for KSat values between 0.2 and 1.7 cm/h and between 2.6 and 3.7 cm/h than the reference category (3.8-22.9 cm/h), respectively, adjusting for all other factors in the model. TC contamination was higher in 2011 and 2012 compared to the reference year, 2010, with the odds of contamination 1.15 times (95% CI 1.03-1.29, p=.01) higher for 2012 and 1.22 (95% CI 1.09-1.37, p=.001) higher for 2011. All months except for May were positively associated with increased TC contamination, with increases between 1.41 (June) (95% CI 1.16-1.72, p=.001) and 2.67 (September) (95% CI 2.21-3.23, p<.001) compared to the reference month (April), adjusting for all other factors in the model.

The Hosmer Lemeshow test of Goodness of Fit based on the decile groups of estimated probability (HL=11.480, df=8, p=.18) indicated that there was no evidence for a lack of fit. The ROC-AOC statistic was 0.62, (95% CI 0.61-0.63, p<.001), indicating that the model has poor predictive efficacy for TC contamination.

Escherichia coli models

Five independent variables were associated in the initial univariable analysis at α = .05 with *E. coli* positivity (Table A.4) including precipitation, density of large animals, KSat, month and year. Because the relationship between precipitation and the logit of the proportion of *E. coli* positivity was linear, precipitation was used as a continuous variable. All variables were associated with *E. coli* positivity at a p<.25, and so were included in the initial multivariate model building process (Hosmer et al. 2013).

The final multivariable *E. coli* model (Table A.5) included the following variables: precipitation, sand, density of large animals, KSat, month and year of collection. No significant interaction terms were identified for inclusion (Table A.6). Precipitation was associated with *E. coli* positivity as a continuous variable in the model. For each increase of 1 mm precipitation, *E. coli* contamination increased 1.01 times (95%CI 1.00-1.01, p<.001), adjusting for all other factors in the model. All categories of percentage of sand in the soil (from 31.4-82.0%) were associated with a between 1.70 and 2.05 times increase in the odds of *E. coli* contamination of the well water sample compared to the reference category (0-30.9%) when adjusting for all other factors in the model. Density of large animals between 15.0 and 25.9 animals/km^2 was associated with a 1.50 times (95% CI 1.02-2.17, p=.04) increase in the odds of *E. coli* contamination of the well water sample compared to the reference category (2.5-14.9 animals/km^2), when adjusting for all other factors in the model. The odds of *E. coli* positivity was between 2.00 and 2.69 higher for soils with a KSat between 0.2 and 3.7 cm/h compared with the reference level (3.8-22.9 cm/hour), adjusting for all other factors in the model. The odds of *E. coli* contamination was 1.45 times (95% CI 1.05-2.00, p=.02) higher in 2011 compared to the reference year, 2012, adjusting for all other factors in the model. The odds of *E. coli* contamination in August was 2.19 times higher than in the reference month, April (95% CI 1.21-3.97, p=.009), and the odds of *E. coli* contamination was 2.02 times higher in September than in the reference month, April (95% CI 1.09-3.75, p=.03), when adjusting for all other factors in the model. No other months were correlated with increased *E. coli* contamination.

The Hosmer Lemeshow test of Goodness of Fit based on the decile groups of estimated probability (HL=4.580, df=8,p=.80) indicated that there was no evidence for a lack of fit. The area under the curve statistic was 0.69 (95% CI 0.65-0.72, p<.001), indicating that the model has acceptable predictive efficacy for *E. coli* contamination (Hosmer et al. 2013).

Discussion

In this study increased TC contamination in rural well water in Central Alberta, Canada was associated with increased zero-till farming, increased sand content, higher density of large animals, decreased saturated hydraulic conductivity, year and month of specimen. In addition, increased *E. coli* contamination in rural well water was associated with increased precipitation, increased sand content, higher density of large animals, decreased saturated hydraulic conductivity, and year and month of specimen.

Extreme precipitation events have been associated with waterborne disease outbreaks across the United States, in Canada, and around the world (Cann et al. 2013; Curriero et al. 2001; Thomas et al. 2006) , while Page et al. (2012) noted increased *E. coli* contamination of both groundwater and surface water during periods of rainfall in an alluvial system. Whereas the relationship between outbreaks and extreme precipitation events has been studied, rarely has this been the case for impact of non-extreme precipitation, as in this study, and water well contamination. This study also demonstrated an association between increased *E. coli* contamination and increased precipitation, though the increase in contamination per mm increase of monthly precipitation was relatively small. While these results are in accordance with results in other studies (Cann et al. 2013; Curriero et al. 2001; Page et al. 2012; Thomas et al. 2006), the collection of the precipitation variable could have been more optimal. Total monthly precipitation was estimated for each quarter section of land sampled in the study for the month of sampling, or the previous month, depending on the day of the month studied. Previous studies have determined a association with a specific lag time of a month or more between extreme precipitation events and waterborne outbreaks, especially with groundwater contamination (Curriero et al. 2001; Rose et al. 2000). However much shorter lag times between precipitation events and increased groundwater bacterial contaminations have been observed. In one study, lag times were greater with a greater distance between a well and surface water (Page et al. 2012). A 2015 study of manure impact on *E. coli* concentrations used precipitation lags of 24, 48, 72, 96 and 120 hours prior to sampling, and found no direct relationship between rainfall and *E. coli* concentration on their particular study site, indicating that rainfall lag and groundwater bacterial concentration may be dependent on a number of other factors (Arnaud et al. 2015). For the State of New Jersey, an optimal lag time of 10 days cumulative precipitation pre-sampling was determined and used in a similar logistic regression study (Procopio et al. 2017). The timing of the influence of precipitation on ground water is complex, and dependent on a number of factors, some specific to individual wells, such as depth, construction, and proximity to surface water. Using a single lag time to predict the influence of precipitation on groundwater contamination for such a large expanse of geography is a compromise. Ideally a large number of different lag times, ranging from 24 hours to more than a month, would be used for each individual well, but because of the samples size of this study, was deemed impractical.

In other studies, bacterial transport into discharge water (tile water) was reduced by tilling the land prior to application of liquid manure (Samarajeewa et al. 2012). Zero-till farming over a number of seasons appeared to have a cumulative effect of increasing the macropore structure in the soil, and was associated with increased volume of subsurface water flow and numbers of *E. coli* in the discharge water (Thiagarajan et al. 2007). Soil columns containing earthworms or decaying roots also facilitated faster transport of *E. coli* compared with packed columns (Safadoust et al. 2011; Steiner 2009). This is, to the knowledge of the authors, the first study showing an association specifically between zero-till farming and increased well water contamination with TC. However, the results should be viewed with caution. The association was only seen between TC and one category of zero till farming, which represented the highest percentage of land devoted to zero-till farming. The effect size was small, an increase in contamination of a factor of 1.3. The same association was not seen between zero-till farming and elevated *E. coli* contamination. It is possible that this result was also due to geographical confounding, as each category of zero-till farming represents results aggregated to a limited number of census regions, and in the case of zero-till farming, the significant category represented two large, contiguous areas of land (Figure 2).

Figure 2: Geographical locations of counties by zero-till categories for the study area in Alberta, Canada.

Animal grazing has been associated with increased *E. coli* levels in runoff water (Muirhead 2009) as well as greater surface water contamination than recreation areas or remote wildlife areas (Derlet et al. 2012). In addition, feces left by grazing sheep under simulated rainfall conditions in New Zealand released *E. coli* in response to rainfall as much as 30 days after deposition (McDowell 2006; Moriarty and Gilpin 2014). An association was found in this study, again only between one level of animal density (15.0-25.9 animals/km^2) and increased *E. coli* contamination. It is difficult to understand how a moderate level of animal density was associated with increased contamination while the highest level of animal density (35.0-44.3 animals/km^2) was not. In addition, an association was found between animal density of 26.3-32.7 animals/km^2 and decreased contamination with TC. These results may suffer from the same problem as identified with the zero-till result, a geographical confounding effect, as well as ecological fallacy, or there may be an impact of different types on animal husbandry between high density and lower density animal farming. Farms with the highest densities of domestic animals may not have any cropland and may sell all manure produced rather than storing or using on site.

The significant association between decreasing KSat values, and increasing *E. coli* contamination, while adjusting for the other factors in the model, is an interesting and somewhat paradoxical result. KSat is a measure of the rate that water moves through saturated soil (Jhonson 2009). The greater the KSat, the more soil nutrients and potentially bacteria will also move through the soil. However, in this study greater KSat was associated with less contamination. This result may be reasonable depending on how KSat was calculated. The soil database specifically states that a large portion of the database was estimated rather than directly measured (Agricultural and Agri-Food Canada 2008). If the KSat values were estimated based on, for instance the sand/clay content in the soil, and macropores or other structural features were not included in the calculation (not specified in the meta data provided by the database, and unlikely given the level of estimation), then the database may only be describing part of the picture. High sand content is associated with high KSat values, because sand has a large average particle size and correspondingly larger pores allow rapid movement of water through the soil. However, high sand content soils are less able to support structure than soils with high clay content (Safadoust et al. 2011). In high clay content soils without macropores, water moves slowly, and larger particles like bacteria are trapped. However, macropores are more likely to form and keep their integrity over time in high clay content soils (Safadoust et al. 2011). Macropores allow water, as well as larger particles like bacteria, to move through the soil quickly (Cey et al. 2009). Therefore, in our study, soil that might be perceived as having a low KSat, and thus be at low risk for bacterial transport, may have secondary pathways for preferential transport of bacteria and, as such, may be at greater risk for bacterial transport than soils with a higher KSat.

Year and month of sample collection were associated with both increased *E. coli* and TC contamination. This is in accordance with the seasonality demonstrated in our previous study using the same body of data (Invik et al. 2017).

The quality of the agricultural variables was a limiting factor in this study. While the Canadian Agricultural Census (Government of Canada 2012) provides a wealth of information, for confidentiality reasons it is aggregated to large regions. Using geographically aggregated data leads to two problems, ecological fallacy and the modifiable areal unit problem (MAUP). Ecological fallacy occurs when results based on aggregated data are erroneously applied to specific members of the aggregated area. If an aggregated area has an average animal density that is high, and animal density is associated with contamination in a study, it is incorrect to infer from that study that well X within that aggregated area will have a high rate of contamination. It is quite possible that the high animal density is based on animal numbers from only one half of the polygon and that well X is in the other half (Dark and Bram 2007). The two key concerns about MAUP are that the decision for the boundaries for the polygons is arbitrary and different decisions on the boundaries will produce different statistical outcomes. In addition, the aggregation of smaller units into larger areas means a decrease in the variation in the data while the mean remains unchanged (Dark and Bram 2007). Data aggregated to smaller regions or point data, such as specific locations of animals' numbers would have improved the quality of the study considerably.

The addition of other variables not currently available to us into the model might have explained the data more thoroughly. Variables that may have been of value included data on individual wells such as depth of well, type of well, condition of well, and aquifer type.

A number of associations were found between TC and *E. coli* contamination and agriculture, soil properties and precipitation variables in this study; however caution should be exercised in interpreting these results. This was a retrospective, observational study of a voluntary testing program. The voluntary nature of the study introduces some interesting potential for bias, which could either inflate the contamination with of *E. coli* and TC or deflate it. For instance, people may be more likely to test when they are concerned about their water quality or when someone in their home is ill, inflating the proportion contaminated. Alternatively, it may be that those who chose to regularly test their wells are also vigilant in ensuring regular maintenance and periodic shock chlorination. It is likely that both factors were at play. The only way to get an idea of the size and direction of bias would be to couple testing of randomly selected wells with a survey of testing attitudes.

Understanding risk factors is not always enough to motivate well owners to test or perform preventative maintenance. A study of septic systems in Ireland looked at risk perceptions and motivations (Devitt et al. 2016). Very much like the survey papers on well owners and risk perception (Summers 2010; Malecki et al. 2017), sensory cues (bad smells) or health concerns were required before people felt the need perform inspections or preventative maintenance. The risk of fines was more highly motivating than worries about environmental or health risks. In addition, fears about replacement costs represented a barrier to septic system owners even inspecting their own systems, in case they were to find an expensive problem brewing. Aside from legislation and fines, some of the suggestions from Devitt et al. (2016) include using positively framed, self-empowering messages , rather than fear inducing messages, use of visual frames of reference and providing accurate information about costs, which may allay some of the fears, as actual costs seemed to often be less than owners feared.

Conclusions

The outcome data covered a large study area and had the advantage of a large number of cases. While these data offers a wealth of information, pairing it with independent variables of high enough resolution to be of value was a difficult task. Three significant factors for increased *E. coli*-positive untreated drinking water wells were identified: soil properties (KSat and sand content), animal density and monthly cumulative precipitation. Awareness of these risk factors and appropriate mitigation strategies such as shock chlorination of wells or more frequent testing of wells when these risk factors are present can help protect users of untreated rural groundwater drinking sources. Messages to well owners empowering them to seek out the benefits of these actions as well as practical information about well water systems and approximate costs of maintenance and replacement may increase motivation for individual owners. Further studies incorporating point data of agricultural variables such as animal numbers and land use practices would be of benefit in further evaluating these risk factors.

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. 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 statistical advice from Henrik Stryhn and Cheryl Waldner and help provided by Peter Peller of the University of Calgary Library with GIS data and software. Thank you to Mathieu Petit for the French translation of the abstract.

References

Agriculture and Agri-Food Canada. 2008. Soil Landscapes of Canada Version 3.2. October 8, 2008. http://sis.agr.gc.ca/cansis/nsdb/slc/v3.2/index.html (accessed June 2018).

Alberta Agriculture and Forestry. 2014. Current and Historical Alberta Weather Station Data Viewer. Current and Historical Alberta Weather Station Data Viewer. 2014. http://agriculture.alberta.ca/acis/alberta-weather-data-viewer.jsp (accessed June 2018).

Alberta Environment and Parks. 2009. Alberta Environment's Drinking Water Program: A 'Source to Tap, Muli-Barrier' Approach. May 2009. http://www.environment.alberta.ca/apps/RegulatedDWQ/More.aspx (accessed June 2018).

Alberta Environment and Parks. 2010. Alberta Township Survey System | AEP - Environment and Parks. Alberta Township Survey System. 2010. http://aep.alberta.ca/recreation-public-use/recreation-on-agricultural-public-land/alberta-township-survey-system.aspx (accessed June 2018).

Arnaud, E., A. Best, B. L. Parker, R. Aravena, and K. Dunfield. 2015. Transport of Escherichia Coli through a Thick Vadose Zone. *Journal of Environmental Quality* 44 (5): 1424–34. https://doi.org/10.2134/jeq2015.02.0067.

Arthur, T. M., T. L. Wheeler, S. D. Shackelford, M. Koohmaraie, X. Nou, J. M. Bosilevac, J. E. Keen, N. Kalchayanand, and D. M. Brichta-Harhay. 2009. Longitudinal Study of *Escherichia Coli* O157:H7 in a Beef Cattle Feedlot and Role of High-Level Shedders in Hide Contamination. *Applied and Environmental Microbiology AEM* 75 (20): 6515–23. https://doi.org/10.1128/AEM.00081-09.

Bergeron, C. R., C. Prussing, L. Dutil, P. Boerlin, D. Daignault, R. J. Reid-Smith, G. G. Zhanel, and A. R. Manges. 2012. Chicken as Reservoir for Extraintestinal Pathogenic *Escherichia Coli* in Humans, Canada. *Emerging Infectious Diseases* 18 (3): 415–21. http://dx.doi.org/10.3201/eid1803.111099.

Berry, E. D., J. E. Wells, T. M. Arthur, B. L. Woodbury, J. A. Nienaber, T. M. Brown-Brandl, and R. A. Eigenberg. 2010. Soil versus Pond Ash Surfacing of Feedlot Pens: Occurrence of *Escherichia Coli* O157:H7 in Cattle and Persistence in Manure. *Journal of Food Protection* 73 (7): 1269–77.

Cann, K. F., D. Rh. Thomas, R. L. Salmon, A. P. Wyn-Jones, and D. Kay. 2013. Extreme Water-Related Weather Events and Waterborne Disease. *Epidemiology & Infection* 141 (4): 671–86. https://doi.org/10.1017/S0950268812001653.

Cey, E. E., D. L. Rudolph, and J. Passmore. 2009. Influence of Macroporosity on Preferential Solute and Colloid Transport in Unsaturated Field Soils. *Journal of Contaminant Hydrology* 107 (1–2): 45–57. https://doi.org/10.1016/j.jconhyd.2009.03.004.

Chapman, P.A., C.A. Siddons, A.T. Gerdan Malo, and M.A. Harkin. 1997. A 1-Year Study of *Escherichia Coli* O157 in Cattle, Sheep, Pigs and Poultry. *Epidemiology and*

Infection 119: 245–50. http://dx.doi.org.ezproxy.lib.ucalgary.ca/10.1017/S0950268897007826.

Curriero, F. C., J. A. Patz, and J. B. Rose. 2001. The Association between Extreme Precipitation and Waterborne Disease Outbreaks in the United States, 1948-1994. *American Journal of Public Health* 91 (8): 1194–99. https://doi.org/10.2105/AJPH.91.8.1194.

Dark, S. J., and D. Bram. 2007. The Modifiable Areal Unit Problem (MAUP) in Physical Geography. *Progress in Physical Geography* 31 (5): 471–79. https://doi.org/10.1177/0309133307083294.

Dassouli-Mrani-Belkebir, A, M Contrepois, J P Girardeau, and M der Vartanian. 1988. Characters of *Escherichia Coli* 078 Isolated from Septicaemic Animals. *Veterinary Microbiology* 17 (4): 345–56.

Deckert, A., S. Gow, L. Rosengren, D. Léger, B. Avery, D. Daignault, L. Dutil, R. Reid-Smith, and R. Irwin. 2010. Canadian Integrated Program for Antimicrobial Resistance Surveillance (CIPARS) Farm Program: Results from Finisher Pig Surveillance. *Zoonoses & Public Health* 57: 71–84. https://doi.org/10.1111/j.1863-2378.2010.01356.x.

Denno, D. M., W. E. Keene, C. M. Hutter, J. K. Koepsell, M. Patnode, D. Flodin-Hursh, L. K. Stewart, et al. 2009. Tri-County Comprehensive Assessment of Risk Factors for Sporadic Reportable Bacterial Enteric Infection in Children. *Journal of Infectious Diseases* 199 (4): 467–76. https://doi.org/10.1086/596555.

Derlet, Robert W., J. R. Richards, L. L. Tanaka, C. Hayden, K. A. Ger, and C. R. Goldman. 2012. Impact of Summer Cattle Grazing on the Sierra Nevada Watershed: Aquatic Algae and Bacteria. *Journal of Environmental & Public Health*, January, 1–7. https://doi.org/10.1155/2012/760108.

Dettinger, M. 2011. Climate Change, Atmospheric Rivers, and Floods in California - A Multimodel Analysis of Storm Frequency and Magnitude Changes. *Journal of the American Water Resources Association* 47 (3): 514–23. https://doi.org/10.1111/j.1752-1688.2011.00546.x.

Devitt, C., E. O'Neill, and R. Waldron. 2016. Drivers and Barriers among Householders to Managing Domestic Waste Water Treatment Systems in the Republic of Ireland; Implications for Risk Prevention Behaviour. *Journal of Hydrology* 535 (April): 534–46. https://doi.org/10.1016/j.jhydrol.2016.02.015.

Evans, S., and L. Li. 2005. A Comparison of Goodness of Fit Tests for the Logistic GEE Model. *Statistics In Medicine* 24 (8): 1245–61.

Garvey, P., A. Carroll, E. McNamara, and P. J. McKeown. 2016. Verotoxigenic Escherichia Coli Transmission in Ireland: A Review of Notified Outbreaks, 2004-2012. *Epidemiology and Infection* 144 (5): 917–26.

Gleeson, C., and N. Gray. 1996. *The Coliform Index and Waterborne Disease: Problems of Microbial Drinking Water Assessment*. London, Great Britain: E & FN Spon.

Government of Alberta. 2014. Water Wells: How to Manage a Water Well. Fact sheet. http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/wqe14939 (accessed June 2018).

Government of Canada. 2012. 2011 Census of Agriculture. 2012. http://www.statcan.gc.ca/eng/ca2011/index accessed June 2018).

Horton, NJ, J D Bebchuk, C L Jones, S R Lipsitz, P J Catalano, G E Zahner, and G M Fitzmaurice. 1999. Goodness-of-Fit for GEE: An Example with Mental Health Service Utilization. *Statistics In Medicine* 18 (2): 213–22.

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. *Wiley Series in Probability and Statistics : Applied Logistic Regression (3rd Edition)*. New York, NY, USA: Wiley.

Hynds, P. D., M. K. Thomas, and K. D. M. Pintar. 2014. Contamination of Groundwater Systems in the US and Canada by Enteric Pathogens, 1990-2013; a Review and Pooled-Analysis. *PloS One* 2014 (e93301). https://doi.org/10.1371/journal.pone.0093301.

IDEXX Laboratories. 2012. Colilert. 2012. https://www.idexx.com/en/water/water-products-services/colilert/ (accessed June 2018).

Invik, J., H. W. Barkema, A. Massolo, N. F. Neumann, and S. Checkley. 2017. Total Coliform and Escherichia Coli Contamination in Rural Well Water: Analysis for Passive Surveillance. *Journal of Water & Health* 15 (5): 729–40. https://doi.org/10.2166/wh.2017.185.

Jhonson, C.. 2009. *Biology of Soil Science*. Oxford Book Co.

Karmali, M. 2002. The Medical Significance of Shiga Toxin-Producing *Eschereichia Coli* Infections. In *E. Coli: Shiga Toxin Methods and Protocols*, edited by D. Philpott and Frank Ebel, 1–7. Totowa, New Jersey: Humana Press.

Krog, J. S., A. Forslund, L. E. Larsen, A. Dalsgaard, J. Kjaer, P. Olsen, and A. C. Schultz. 2017. Leaching of Viruses and Other Microorganisms Naturally Occurring in Pig Slurry to Tile Drains on a Well-Structured Loamy Field in Denmark. *Hydrogeology Journal* 25 (4): 1045–62. https://doi.org/10.1007/s10040-016-1530-8.

Liu X., Z. YuQin, Y. Meng, Z. HongYun, W. Peng, Z. WenJuan, and Y. Hui. 2015. The Epidemiological Characteristics Analysis and the Pathogen Source Tracking of a Bacterial Dysentery Outbreak. *Modern Preventive Medicine* 42 (16): 2904–7.

Malecki, K. M. C., A. A. Schultz, D. J. Severtson, H. A. Anderson, and J. A. VanDerslice. 2017. Private-Well Stewardship among a General Population Based Sample of Private Well-Owners. *Science of the Total Environment* 601/602: 1533–43.

McDowell, R. W. 2006. Contaminant Losses in Overland Flow from Cattle, Deer and Sheep Dung. *Water, Air & Soil Pollution* 174 (1–4): 211–22. https://doi.org/10.1007/s11270-006-9098-x.

Michel, P., J. B. Wilson, S. W. Martin, R. C. Clarke, S. A. McEwen, and C. L. Gyles. 1999. Temporal and Geographical Distributions of Reported Cases of *Escherichia Coli* O157:H7 Infection in Ontario. *Epidemiology and Infection* 122 (2): 193–200. https://doi.org/10.1017/S0950268899002083.

Mitchell, J. P., P. N. Singh, W. W. Wallender, D. S. Munk, J. F. Wroble, W. R. Horwath, P. Hogan, R. Roy, and B. R. Hanson. 2012. No-Tillage and High-Residue Practices Reduce Soil Water Evaporation. *California Agriculture* 66 (2): 55–61. https://doi.org/10.3733/ca.v066n02p55.

Moriarty, EM, and BJ Gilpin. 2014. Leaching of *Escherichia Coli* from Sheep Faeces during Simulated Rainfall Events. *Letters in Applied Microbiology* 58 (6): 569–75. https://doi.org/10.1111/lam.12230.

Muirhead, R. W. 2009. Soil and Faecal Material Reservoirs of *Escherichia Coli* in a Grazed Pasture. *New Zealand Journal of Agricultural Research* 52 (1): 1–8.

Mukhi, S. N., J. May-Hadford, S. Plitt, J. K. Preiksaitis, and B. E. Lee. 2011. DIAL: A Platform for Real-Time Laboratory Surveillance. *Online Journal of Public Health Informatics* 2 (3). https://doi.org/10.5210/ojphi.v2i3.3041.

Murphy, H. M., M. K. Thomas, P. J. Schmidt, D. T. Medeiros, S. McFadyen, and K. D. M. Pintar. 2016. Estimating the Burden of Acute Gastrointestinal Illness Due to Giardia, Cryptosporidium, Campylobacter, E. Coli O157 and Norovirus Associated with Private Wells and Small Water Systems in Canada. *Epidemiology and Infection* 144 (7): 1355–70. https://doi.org/10.1017/S0950268815002071.

Neumann, N. F, D. W Smith, and M. Belozevic. 2005. Waterborne Disease: An Old Foe Re-Emerging? *Journal of Environmental Engineering & Science* 4 (3): 155–71. https://doi.org/10.1139/S04-061.

Page, R., S. Scheidler, E. Polat, P. Svoboda, and P. Huggenberger. 2012. Faecal Indicator Bacteria: Groundwater Dynamics and Transport Following Precipitation and River Water Infiltration. *Water, Air & Soil Pollution* 223 (5): 2771–82. https://doi.org/10.1007/s11270-011-1065-5.

Pan, W. 2001. Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics* 57 (1): 120–25.

Procopio, N. A., T. B. Atherholt, S. M. Goodrow, and L. A. Lester. 2017. The Likelihood of Coliform Bacteria in NJ Domestic Wells Based on Precipitation and Other Factors. *Ground Water* 55 (5): 722–35.

Rose, J. B., S. Daeschner, D. R. Easterling, F. C. Curriero, S. Lele, and J. A. Patz. 2000. Climate and Waterborne Disease Outbreaks. *Journal American Water Works Association* 92 (9): 77–87.

Safadoust, A., A.A. Mahboubi, B. Gharabaghi, M.R. Mosaddeghi, P. Voroney, A. Unc, and Gh. Sayyad. 2011. Bacterial Filtration Rates in Repacked and Weathered Soil Columns. *Geoderma* 167–168 (November): 204–13. https://doi.org/10.1016/j.geoderma.2011.08.014.

Samarajeewa, A D, S M Glasauer, J D Lauzon, I P O'Halloran, Gary W Parkin, and K E Dunfield. 2012. Bacterial Contamination of Tile Drainage Water and Shallow Groundwater under Different Application Methods of Liquid Swine Manure. *Canadian Journal Of Microbiology* 58 (5): 668–77.

Scaife, A., T. Spangehl, D. Fereday, U. Cubasch, U. Langematz, H. Akiyoshi, S. Bekki, et al. 2012. Climate Change Projections and Stratosphere-Troposphere Interaction. *Climate Dynamics* 38 (9/10): 2089–97. https://doi.org/10.1007/s00382-011-1080-7.

Steiner, P. N.. 2009. Effects of Growing and Desiccated Roots or *E. Coli* Movement through Soil Columns. Masters of Science, USA: Michigan State University.

Summers, R. 2010. Alberta Water Well Survey: A Report Prepared for Alberta Environment. http://aep.alberta.ca/water/programs-andservices/groundwater/documents/AlbertaWaterWellSurvey-Report-Dec2010.pdf (accessed June 2018).

Technical Advisory Committee on Safe Drinking Water. 2007. Environmental Public Health Field Manual for Private, Public and Communal Drinking Water Systems in Alberta. Third Edition. Alberta Health and Wellness.

Thiagarajan, A., R. Gordon, A. Madani, and G. W. Stratton. 2007. Discharge of *Escherichia Coli* from Agricultural Surface and Subsurface Drainage Water: Tillage Effects. *Water*, no. 1: 3–12. https://doi.org/10.1007/s11270-006-9300-1.

Thomas, MK, DF Charron, D Waltner-Toews, C Schuster, AR Maarouf, and JD Holt. 2006. A Role of High Impact Weather Events in Waterborne Disease Outbreaks in Canada, 1975 - 2001. *International Journal of Environmental Health Research* 16 (3): 167–80.

Valcour, J. E., P. Michel, S. A. McEwen, and J. B. Wilson. 2002. Associations between Indicators of Livestock Farming Intensity and Incidence of Human Shiga Toxin-Producing *Escherichia Coli* Infection. *Emerging Infectious Diseases* 8 (3): 252–57. https://doi.org/10.3201/eid0803.010159.

Wallender, E. K., E. C. Ailes, J. S. Yoder, V. A. Roberts, and J. M. Brunkard. 2014. Contributing Factors to Disease Outbreaks Associated with Untreated Groundwater. *Ground Water* 52 (6): 886–97.

World Health Organization. 2018. Drinking-Water. WHO. February 2018. http://www.who.int/en/news-room/fact-sheets/detail/drinking-water (accessed June 2018).

World Health Organization. 2011. Guidelines for Drinking-Water Quality, Fourth Edition 2011. WHO. 2011. http://www.who.int/water_sanitation_health/water-quality/guidelines/dwq-guidelines-4/en/ (accessed June 2018).

Zeger, S L, and K Y Liang. 1986. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics* 42 (1): 121–30.

Appendix A

Table A.1 Unconditional associations of various predictor variables with total coliform positive well water samples using logistic regression with generalized estimating equations for 2010-2012 (April-Oct), Alberta, Canada.

Variable	Odds Ratio	OR 95% CI	Wald Chi Square	df	p-value
Precipitation Overall			35.019	4	**<.001**
Precipitation 99.05-230.77 mm/month	1.22	1.07-1.39	9.074	1	**.003**
Precipitation 64.56-99.04 mm/month	0.86	0.75-0.99	4.719	1	**.03**
Precipitation 45.77-64.55 mm/month	0.88	0.77-1.00	4.017	1	**.04**
Precipitation 24.47-45.76 mm/month	0.99	0.87-1.130	0.015	1	.90
Precipitation 0.877-24.461 mm/month	Ref				
Zero Till Overall			10.789	3	**.01**
Zero Till 24.1-61.9%	1.19	0.96-1.47	2.452	1	.12
Zero Till 17.7-21.8%	0.90	0.71-1.14	0.770	1	.38
Zero Till 6.0-17.2%	1.13	0.91-1.41	1.254	1	.26
Zero Till 0-5.1%	Ref				
Sand Overall			7.648	3	.05

Sand 42.9-81.9%	1.15	0.97-1.40	2.483	1	.12
Sand 38.8-42.8%	1.16	0.94-1.45	1.845	1	.17
Sand 31.4-38.7%	1.25	0.94-1.47	7.627	1	**.006**
Sand 0-30.9%	Ref				
Density of large animals Overall			11.329	3	**.01**
Density of large animals 35.0-44.3 animals/km^2	0.86	0.70-1.05	2.188	1	.14
Density of large animals 26.3-32.7 animals/km^2	0.83	0.68-1.03	2.923	1	.09
Density of large animals 15.0-25.9 animals/km^2	1.07	0.87-1.31	0.402	1	.53
Density of large animals 2.5-15.0 animals/km^2	Ref				
KSat Overall			7.290	3	.06
KSat .2-1.7 cm/hour	1.30	1.06-1.59	6.302	1	**.01**
KSat 1.7-2.5 cm/hour	1.21	0.99-1.49	3.324	1	.07
KSat 2.6-3.7 cm/hour	1.28	1.05-1.57	5.974	1	**.02**
KSat 3.8-22.9 cm/hour	Ref				
Year overall			21.419	2	**<.001**
2012	1.17	1.05-1.31	8.252	1	**.004**

2011	1.27	1.15-1.41	21.087	1	**<.001**
2010	Ref				
Month overall			267.872	6	**<.001**
May	1.07	0.80-1.30	0.443	1	.51
Jun	1.37	1.14-1.70	11.500	1	**.001**
July	2.12	1.84-2.65	72.999	1	**<.001**
Aug	2.50	2.08-3.00	98.875	1	**<.001**
Sept	2.56	2.13-3.08	100.169	1	**<.001**
Oct	2.37	1.96-2.87	78.753	1	**<.001**
Apr	Ref				

Table A.2 Final multivariable model for total coliform contamination for 2010-2012 (April-Oct), Alberta, Canada.

Variable	Odds Ratio	95% CI OR	Wald Chi-Square	df	p-value
Precipitation Overall			10.714	4	**.03**
Precipitation 99.1-230.8 mm/month	1.13	.95-1.34	1.945	1	.163
Precipitation 64.6-99.0 mm/month	0.93	0.78-1.10	0.789	1	.37
Precipitation 45.8-64.5 mm/month	0.98	0.84-1.15	0.059	1	.81
Precipitation 24.5-45.8 mm/month	1.10	0.94-1.26	1.407	1	.24
Precipitation .087-24.5 mm/month	Ref				

Zero Till Overall			9.987	3	**.02**
Zero Till 24.1-61.9%	1.29	1.05-1.59	5.590	1	**.02**
Zero Till 17.7-21.8%	1.02	0.78-1.33	0.020	1	.88
Zero Till 6.0-17.2%	1.18	0.95-1.46	2.220	1	.14
Zero Till 0-5.1%	Ref				
Sand Overall			11.006	3	**.01**
Sand 42.9-81.9%	1.38	1.09-1.75	7.174	1	**.007**
Sand 38.8-42.8%	1.22	1.01-1.48	4.0514	1	**.04**
Sand 31.4-38.7%	1.30	1.10-1.55	8.423	1	**.004**
Sand 0-30.9%	Ref				
Animal Density Overall			10.668	3	**.01**
Density of large animals 35.0-44.3 animals/km^2	0.91	0.72-1.13	0.751	1	.39
Density of large animals 26.3-32.7 animals/km^2	0.77	0.63-0.93	6.941	1	**.008**
Density of large animals 15.0-25.9 animals/km^2	1.02	0.846-1.24	0.061	1	.81
Density of large animals 2.5-15.0 animals/km^2	Ref				

KSat Overall			9.458	3	**.02**
KSat .2-1.7 cm/hour	1.48	1.18-1.91	9.123	1	**.003**
KSat 1.7-2.5 cm/hour	1.31	1.01-1.71	4.019	1	.05
KSat 2.6-3.7 cm/hour	1.31	1.05-1.64	5.620	1	**.02**
KSat 3.8-22.9 cm/hour	Ref				
Year Overall			12.272	2	**.002**
2012	1.15	1.03-1.29	6.338	1	**.01**
2011	1.22	1.09-1.37	11.559	1	**.001**
2010	Ref				
Month Overall			259.746	6	**<.001**
May	1.09	0.90-1.33	0.676	1	.41
Jun	1.41	1.16-1.72	11.688	1	**.001**
Jul	2.17	1.76-2.68	51.976	1	**<.001**
Aug	2.51	2.05-3.05	82.482	1	**<.001**
Sep	2.67	2.21-3.23	104.465	1	**<.001**
Oct	2.42	2.00-2.93	79.018	1	**<.001**
Apr	Ref				

Table A.3 Interaction terms tested in the total coliform model and their significances:

Interaction term	Significance
Zero Till*Animal Density	.606
Zero Till*KSat	.079
Zero Till*Precipitation	.907
Animal Density*KSat	.087
Animal Density*Precipitation	.410
KSat*Precipitation	.589
Month*Animal Density	.287
Month*Zero Till	.109
Month*KSat	.161

Table A.4 Unconditional associations of various predictor variables with *E. coli* positive well water samples using logistic regression with generalized estimating equations for 2010-2012 (April-Oct), Alberta, Canada.

Variable	Odds Ratio	95% CI OR	Wald Chi-Square	Df	p-value
Precipitation	1.01	1.01-1.01	39.131	1	**<.001**

Zero Till Overall			7.189	3	.07
Zero Till 24.1-61.9%	1.36	0.89-2.07	2.061	1	.15
Zero Till 17.7-21.8%	0.82	0.50-1.33	0.661	1	.42
Zero Till 6.0-17.2%	1.24	0.81-1.91	0.985	1	.32
Zero Till 0-5.1%	Ref				
Sand Overall			7.749	3	.05
Sand 42.9-81.9%	1.38	0.91-2.10	2.342	1	.13
Sand 38.8-42.8%	1.62	1.06-2.47	5.059	1	**.02**
Sand 31.4-38.7%	1.70	1.14-2.52	6.856	1	**.009**
Sand 0-30.9%	Ref				
Animal Density Overall			16.069	3	**.001**
Density of large animals 35.0-44.3 animals/km^2	0.86	0.58-1.29	0.527	1	.47
Density of large animals 26.3-32.7 animals/km^2	0.67	0.04-1.02	3.467	1	.06
Density of large animals 15.0-25.9 animals/km^2	1.42	0.97-2.09	3.183	1	.07
Density of large animals	Ref				

2.5-14.9 animals/km^2					
KSat Overall			9.586	3	**.02**
KSat .2-1.7 cm/hour	1.90	1.25-2.89	8.877	1	**.003**
KSat 1.7-2.5 cm/hour	1.65	1.07-2.55	5.028	1	**.03**
KSat 2.6-3.7 cm/hour	1.75	1.15-2.68	6.694	1	**.01**
KSat 3.8-22.9 cm/hour	Ref				
Year Overall			7.066	2	**.03**
2010	1.10	0.79-1.34	0.292	1	.59
2011	1.48	1.08-2.04	5.819	1	**.02**
2012	Ref				
Month Overall			47.796	6	**<.001**
May	0.74	0.36-1.52	0.660	1	.42
Jun	1.70	0.94-3.08	3.030	1	.08
Jul	3.09	1.80-5.41	15.683	1	**<.001**
Aug	3.13	1.78-5.50	15.702	1	**<.001**
Sep	2.07	1.12-3.84	5.388	1	**.02**
Oct	1.69	0.88-3.26	2.486	1	.12
Apr	Ref				

Table A.5 Final multivariable model for *Escherichia coli* contamination for 2010-2012 (April-Oct), Alberta, Canada.

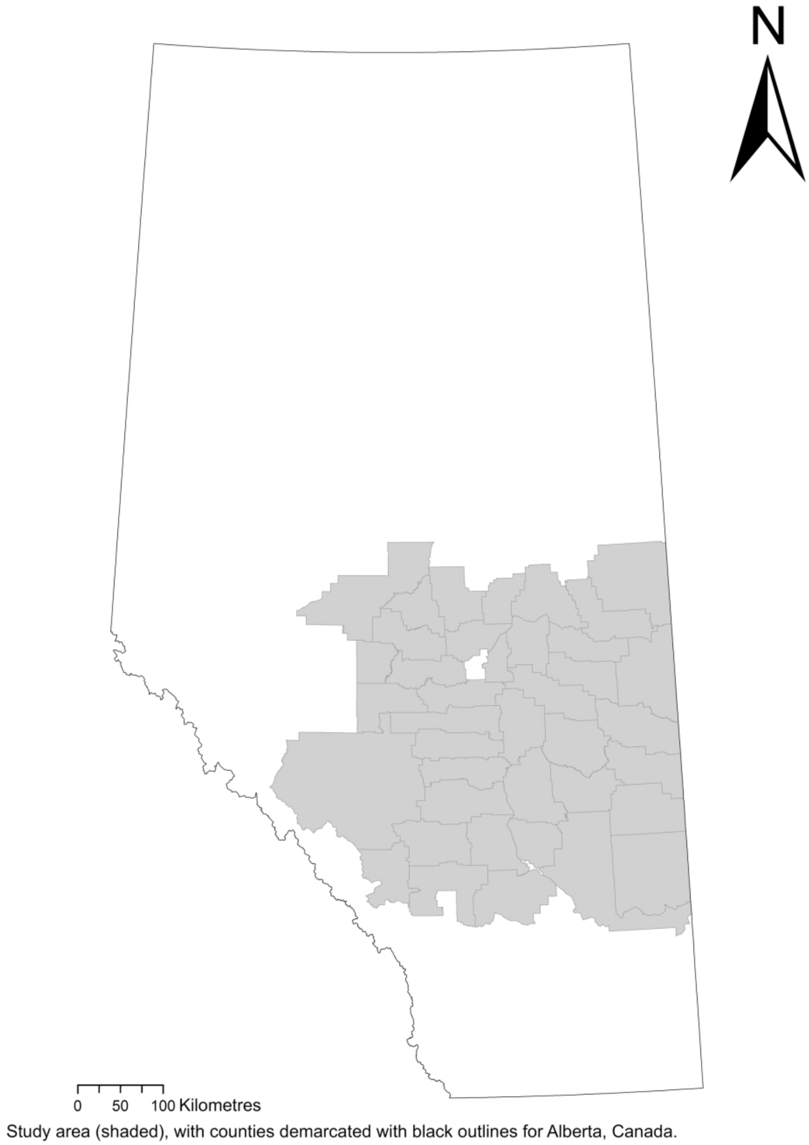
Variable	Odds Ratio	95% CI OR	Wald Chi-Square	df	p-value
Precipitation	1.007	1.003-1.01	15.453	1	**<.001**
Sand Overall			9.332	3	**.03**
Sand 42.9-81.9%	2.05	1.19-3.53	6.771	1	**.009**
Sand 38.8-42.8%	1.86	1.21-2.87	7.840	1	**.005**
Sand 31.4-38.7%	1.70	1.07-2.70	5.123	1	**.02**
Sand 0-30.9%	Ref				
Animal Density Overall			15.732	3	**.001**
Density of large animals 35.0-44.3 animals/km^2	1.00	0.66-1.46	0.011	1	.92
Density of large animals 26.3-32.7 animals/km^2	0.67	0.44-1.01	3.604	1	.06
Density of large animals 15.0-25.9 animals/km^2	1.48	1.02-2.16	4.320	1	**.04**
Density of large animals 2.5-15.0 animals/km^2	Ref				
KSat overall			14.698	3	**.002**
KSat .2-1.7 cm/hour	2.69	1.60-4.50	14.153	1	**<.001**

KSat 1.7-2.5 cm/hour	2.13	1.23-3.70	7.241	1	**.007**
KSat 2.6-3.7 cm/hour	2.00	1.27-3.13	9.078	1	**.003**
KSat 3.8-22.9 cm/hour					
Year Overall			6.141	2	.05
Year 2010	1.10	0.78-1.53	0.283	1	.59
Year 2011	1.45	1.05-2.00	5.076	1	**.02**
Year 2012	Ref				
Month Overall			25.266	6	**<.001**
May	0.67	0.33-1.36	1.239	1	.27
Jun	1.25	0.67-2.33	0.506	1	.48
Jul	1.82	0.99-3.33	3.722	1	.05
Aug	2.19	1.21-3.97	6.772	1	**.009**
Sep	2.02	1.09-3.75	4.955	1	**.03**
Oct	1.79	0.92-3.47	2.978	1	.08
Apr	Ref				

Table A.6 Interaction terms tested in the *Escherichia coli* model and their significances:

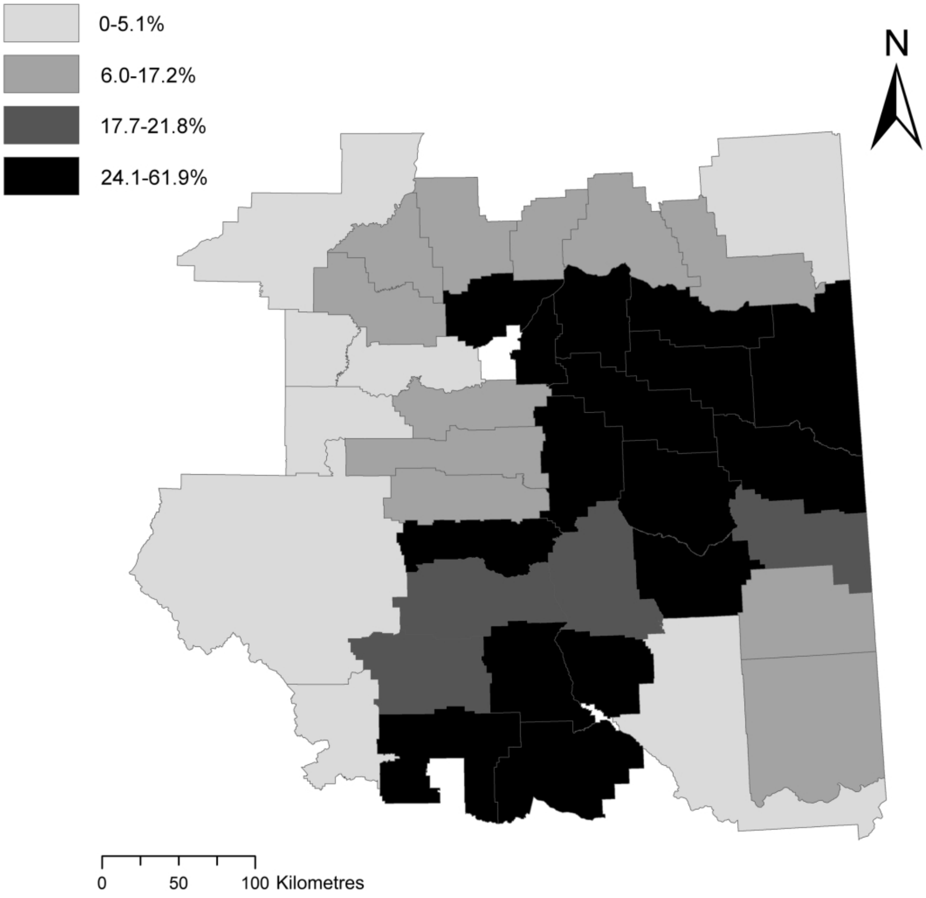
Interaction term	Significance
Sand*Animal Density	.296
Sand *KSat	Quasi-separation, analysis not completed

Sand*Precipitation	.543
Animal Density * KSat	.065
Animal Density * Precipitation	.415
KSat*Precipitation	.727



Study area (shaded), with counties demarcated with black outlines for Alberta, Canada.

215x279mm (300 x 300 DPI)



Geographical locations of counties by zero-till categories for the study area in Alberta, Canada.

Geographical locations of counties by zero-till categories for the study area in Alberta, Canada.

203x198mm (300 x 300 DPI)