Environmental Risk Factors Associated with Contamination of Private Drinking Water Wells with *Escherichia coli*

by

Marie Beliveau

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science
in
Epidemiology

Department of Public Health Sciences

University of Alberta

Abstract

The focus of this research was to identify environmental risk factors that play a critical role in the contamination of drinking water wells with *Escherichia coli*. This study used data from a case control-study carried out between April 2005 and September 2006 in Ontario and Alberta, using private water samples submitted from participating public health laboratories. Significant risk factors associated with *Escherichia coli* contamination of private drinking water wells included the type of water well used, the age of the water well, the housing of livestock on the property and total household income. The age of the water well was identified as a mediator for the type of water well used. The associations between environmental risk factors and *Escherichia coli* contamination of private drinking water wells provide rational for public health programs to highlight factors home owners should consider when working to prevent well water contamination and identify potential sources of contamination.

Acknowledgement

I would like to express the deepest appreciation to my committee co-chairs, Dr. Brenda Coleman and Dr. Norman Neumann. Without their supervision and constant help, this thesis would not have been possible.

I would also like to thank my other committee member, Dr. Ambikaipakan Senthilselvan, for his invaluable help during the analysis of my data.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AU	animal unit
CI	confidence interval
km	kilometre
m	metre
mm	milimetre
OR	odds ratio

CHAPTER 1

Environmental Risk Factors Associated with Contamination of Drinking Water

1.1 Background

The prevention of enteric illness remains an important public health concern in North America (Charrois, 2010; Denno *et al.*, 2009). In Canada, laboratory confirmed cases of many known enteric pathogens are reportable at a provincial, territorial and national level (Schuster *et al.*, 2005). The majority of these reported illnesses are sporadic in nature and not associated with a particular outbreak, making the identification of a source difficult (Denno *et al.*, 2009). Further complicating this issue is the knowledge that these diseases are widely under-reported as a result of their self-limiting nature (Schuster *et al.*, 2005). While the transmission of enteric illnesses through drinking water has been well established (Reynolds *et al.*, 2008; Leclerc *et al.*, 2002), studies into routes of transmission continue to point to drinking water wells as potential sources of enteric illnesses. Two recent case-control studies in the United States identified the drinking of well water as a risk factor for children's enteric infections (Gorelick *et al.*, 2011; Denno *et al.*, 2009). Worldwide, in 2001 it was estimated that 20% of the population lacked access to safe drinking water, and more than 5 million people died annually from illnesses associated with microbiologically contaminated drinking water (Thomas *et al.*, 2006, Hunter *et al.*, 2001).

In the United States, most waterborne outbreaks are associated with non-community water systems, mainly private or communal ground water wells (Aramini *et al.*, 2000; Committee on Environmental Heath and Committee on Infectious Disease, 2009; MMWR, 1999). Similarly, out of 288 outbreaks related to drinking water in

Canada reported between 1974 and 2001, the majority occurred in semi-public systems which are privately owned systems used to provide drinking water to the visiting public, and private systems which are privately owned drinking water sources (Schuster *et al.*, 2005). Reynolds *et al.* (2008) reported that there are more than 140 known microorganisms recognized as waterborne pathogens. These microorganisms can be divided into three main types, bacteria, viruses and protozoa (Health Canada, 2012). Epidemiological investigations into numerous enteric outbreaks associated with *Campylobacter spp., Escherichia coli, Salmonella spp., Cryptosporidium spp.* and *Giardia spp,* have implicated drinking water as the source of infection (Mac Kenzie *et al.*, 1994; Pebody *et al.*, 1999; Olsen *et al.*, 2002; MMWR, 1999; Gallay *et al.*, 2006; Martin *et al.*, 2006; O'Reilly *et al.*, 2007; Jakopanec *et al.*, 2008). Both Canada and the United States have reported large outbreaks of *E. coli* O157:H7 and *Cryptosporidium* associated with the consumption of contaminated drinking water and which have received considerable media attention (Mac Kenzie *et al.*, 1994; Stirling *et al.*, 2001; Schuster *et al.*, 2005; Bigras-Poulin *et al.*, 2004). This includes a large outbreak of *E. coli* in Walkerton, Ontario in 2000 which resulted in over 2000 illnesses and 7 deaths associated with the consumption of contaminated municipal well water (Bigras-Poulin *et al.*, 2004) and an outbreak of *Cryptosporidium* in North Battleford, Saskatchewan in which it was estimated over 5000 cases developed gastroenteritis as a result of the outbreak (Stirling *et al.*, 2001).

While more difficult to assess, association between drinking water and sporadic cases of enteric disease has also been shown (Uhlmann *et al.*, 2009). Pathogens associated with sporadic waterborne illness include *Campylobacter* (Nygard *et al.*, 2004), verotoxigenic *Escherichia coli* (Chalmers *et al.*, 2000), *Salmonella* (Kapperud *et al.*, 1998) and *Cryptosporidium* and *Giardia* (Isaac-Renton *et al.*, 1999). The fact that waterborne illness continues to be such concern in developed countries like Canada, given our access to proven methods for providing safe drinking water, is concerning.

The health effects of pathogenic bacteria, viruses and protozoa in drinking water vary. The most common health effect is gastrointestinal illness. Bacteria such as *Shigella spp* and *Campylobacter spp*, viruses such as norovirus and Hepatitis A and protozoa such as *Giardia spp* and *Cryptosporidium spp* can be responsible for severe gastrointestinal illness; others may infect an individual's lungs, skin, eyes, central nervous system or liver (Health Canada, 2012). In addition to the morbidity caused by initial infection with these organisms, longer lasting health effects may occur. In the case of *E. coli* O157:H7, while most infections are self-limiting and resolve within a week, it can progress to haemorrhagic colitis (HC) and 5 to 10% of these cases, in turn, progress to haemolytic uraemic syndrome (HUS), a life-threatening complication with significant morbidity in survivors, especially in children under five years of age and the elderly (Yoon *et al.*, 2008; Yuling *et al.*, 2000). In addition to potential illnesses, water contaminated with *E. coli* that are resistant to antibiotics has also been associated with the carriage of resistant *E. coli* in humans (Coleman *et al.*, 2012). This raises concerns about the role of drinking water in the proliferation of antimicrobial resistance. Especially when considering research such as the registry-based cohort study carried out by Helms *et al.* (2005) into the risk of invasive illness and death associated with infection with antimicrobial resistant *Campylobacter* strains. Helms *et al.* (2005) found an increased risk of invasive illness and death in cases infected with antimicrobial resistant *Campylobacter* strains compared to patients infected with susceptible strains.

The microbiological examination of water is used worldwide to monitor and control for the quality and safety of various uses, including water used for drinking. Currently available detection methods do not allow for routine analysis of all microorganisms that could be present (Health Canada, 2012) and direct monitoring of waterborne human pathogens is seen as impractical due to low concentrations of the organisms in the water, a wide variety of targets and high cost of laboratory analysis (Savichtcheva and Okabe, 2006). Given that most waterborne outbreaks are related to fecal pollution of water sources, microbiological testing is largely based on the need to identify indicators of fecal pollution (Barrell *et al.*, 2000). In Canada, microbiological water safety and quality in private drinking water wells are assessed by screening for total coliforms, which are used as indicators of the general condition of the well and the plumbing system and *Escherichia coli*, which is used as an indicator of the presence of fecally-derived bacteria and consequently the potential for enteric pathogens to be present (Health Canada, 2012).

Escherichia coli are gram-negative bacteria belonging to the Enterobacteriaceae family (Naylor *et al.*, 2005). For the most part, *E. coli* live as a commensal organism and are considered a normal part of the microbiota of the lower gastro-intestinal tract of mammals, including humans (Elena *et al.*, 2005; Naylor *et al.* 2005). Health Canada's *Guidelines for Canadian Drinking Water Quality* state that, "The maximum acceptable concentration of *E. coli* in public, semi-public and private drinking water systems is none detected in 100ml" (Health Canada, 2012), reflecting a zero tolerance for the presence of this bacteria in drinking water supplies of any kind, including raw drinking water.

Groundwater sources for drinking water can be supplied through a number of well types. Dug wells, which are shallow holes that are dug approximately 3 to 9 meters deep, driven or sandpoint wells, where pipe is driven through gravel or sandy soil approximately 15 meters deep, and drilled wells, which are typically 30 to 120 meters deep, and reach bedrock (Committee on Environmental Heath and Committee on Infectious Disease, 2009). In addition, shore wells or spring wells may also be used. These wells use surface water bodies or shallow groundwater sources (New York State Department of Health, 2008). In areas where the water supply is provided as part of the municipal services, the ongoing maintenance and operation are conducted by a public or private utility that must comply with local legislative requirements (Simpson, 2004). When it comes to private water wells, this responsibility falls to the well owner. The result is that private water wells may not be properly maintained while in use and are thus at an increased risk for contamination of the water supply (Richardson *et al.*, 2009; Simpson, 2004; Corkal *et al.*, 2004; Ritter *et al.*, 2002; Audette *et al.*, 2001). In the United States, the proportion of annual waterborne disease outbreaks associated with non-community ground water supply systems increased between 1976 and 2006 relative to the total number of outbreaks reported in all system types (Craun *et al.*, 2010). While the risk of waterborne outbreaks may be decreasing in public systems, a similar decrease has not been noted in private and semi-private drinking water systems. Research into private well water quality across Canada has shown that between 20 and 40 percent of private water wells fall outside of safe drinking water guidelines set in Health Canada's

Guidelines for Canadian Drinking Water Quality (Summers, 2010). In 2006, Environment Canada reported that groundwater served as the primary source of drinking water for approximately 30% of Canadians, and approximately 4 million of these groundwater users reportedly lived in rural areas and accessed private supplies (Environment Canada, 2006). While the population of private well water users is large, it is fragmented across the country, making it difficult to quantify and monitor these systems, and hindering efforts to educate the operators about well water quality (Charrois, 2010). Based on the 2010 Alberta Water Well Survey, 400 000 to 450 000 Albertans rely on privately owned water wells for their supply of drinking water (Summers, 2010). Similarly, in Ontario in 2009 approximately 1.1 million residents accessed non-municipal drinking water supplies including private drinking water wells and surface water (Statistics Canada, 2009).

Investigations to identify the routes of drinking water contamination are necessary to better understand the role of microbial transmission from the environment and to help focus the implementation of effective preventive measures. Given that enteric diseases are widely under-reported, with a Canadian study estimating that over 300 cases of infectious acute gastrointestinal illness occur in the community for every laboratory-confirmed case reported to provincial public health authorities, the true scope of this problem is likely very large (Schuster *et al.*, 2005). The Public Health Agency of Canada estimates that 1 in 8 Canadians become ill as a result of domestically acquired food-borne diseases, which would include illness that results from the consumption of contaminated drinking water (Thomas *et al.*, 2013).

Compared to the foodborne route of transmission, waterborne transmission of enteropathogenic microorganisms can be more difficult to study and specific control measures harder to implement (Bigras-Poulin *et al.*, 2004). This is likely because several factors typically contribute to an outbreak. For example, in the case of the *E. coli* O157:H7 outbreak in Walkerton, Ontario, heavy rainfall, an insecure well head, the presence of pathogenic bacteria in the environment, inadequate water treatment and human error all played a role in disease outbreak (Schuster *et al.*, 2005). A better understanding of the environmental risk factors that play a role in the contamination of private drinking water wells will help to inform public health policy decisions on the development of preventive programs to help reduce the incidence of enteric illness related to private drinking water.

1.2 Review of Environmental Risk Factors Associated with Groundwater Quality

Environmental risk factors for the contamination of groundwater used as a private drinking water source can be divided into three main categories:

1) *Physical Factors*: Including the depth and type of well used, soil type and proximity to potential point-sources of contamination such as private septic systems and manure storage.
2) *Land Use Factors*: Including rural versus urban settings, population density, and proximity to non-point sources such as the number and type of livestock that may be housed on the property.
3) *Temporal Factors*: Including seasonal variation in the well water quality and climatic data including rainfall, daily temperatures and overland flooding.

1.2.1 Physical Factors

A variety of physical factors can affect the vulnerability of a private drinking water well to bacteriological contamination including the location, construction and proximity of the well to sources of pollution (Richardson *et al.*, 2009). A study carried out by Richardson *et al.* (2009) on the microbial water quality of private water supplies in England found an association between the probability of failure (*E. coli*>1/100mL) and the water supply type, whether the water was treated or not and the source of the water. They found that private domestic wells were more likely to fail than commercial ones, that water supplies with no effective treatment were more likely to fail than those that were treated and that water taken from spring wells and surface water was more likely to fail than drilled ground water wells. These findings are expected given the requirement for commercial water sources to comply with legislation intended to protect drinking water quality, and given the quality of groundwater from drilled wells, when properly maintained, is at a lower risk of direct fecal contamination as compared to surface waters which are more easily influenced by fecal sources of pollution (i.e., wildlife, agricultural animals, human sewage) and the environmental factors that transport these sources directly to surface water.

A study on contamination in Ontario farmstead domestic wells carried out by Goss *et al.* (1998) found that well depth was a particularly important variable associated with contamination. At well depths between 10 and 20 meters, a decline in contamination was noted when compared to shallower depths. Similarly, a study carried out by Audette *et al.* (2001) on farm well water quality in Alberta found that shallow wells, which were classified as less than 30 meters deep, were more likely to be contaminated with coliform bacteria than wells greater than 30 meters deep. Goss *et al.* (1998) also determined that the combination of well type and depth accounted for the largest variation in contamination rates, with dug or bored wells, which are typically less than 15 meters deep, being more frequently associated with contamination than drilled wells, which are typically more than 30 meters deep. In addition, Goss *et al.* (1998) found a statistically significant increase in the concentration of bacteria as the age of the well increased. However, the effect of age decreased with well depth, and at a depth of 30 meters, the frequency of contamination in wells 100 years old was only 5% greater than wells that were 5 years old. As a result, the impact of age on drilled wells, versus dug or bored wells, was generally small because of the interaction with the well's depth.

Managing household wastewater is also an issue that affects the majority of rural residents nation-wide. On-site private wastewater treatment is commonly dealt with using an in-ground trench system that is comprised of a septic tank coupled with a soil absorption system (Mathis *et al.*, 2011). Wastewater, which includes both grey water (water from sources such as a washing machine or shower) and black water (water from toilets), flows from the home into a tank where solids settle to the bottom and fats, grease and other floatable material create a top layer (Mathis *et al.*, 2011). The wastewater effluent, the middle layer, is then distributed to subsurface trenches that make up the filter field where it is eventually reincorporated into the natural water cycle (Mathis *et al.*, 2011). Other options for dealing with wastewater include surface discharge as well as the use of pits, ponds or lagoons (Shook *et al.*, 1993). The risk of ground water contamination from private wastewater systems that consist of surface discharge or pits and lagoons is significantly higher than the use of a septic tank with a leaching field

(Shook *et al.*, 1993). The likelihood that a septic tank with a leaching field will become a risk factor to groundwater contamination relates primarily to installation of a system: i.e., correctly identifying the soil conditions in the area and ensuring proper maintenance of that system once in place (Mathis *et al.*, 2011). While malfunctioning or poorly maintained septic systems have been the cause of outbreaks of waterborne disease in the United States attributed to contaminated ground water (Ritter *et al.*, 2002; Borchardt *et al.*, 2003), Goss *et al.* (1998) found no significant effects on the level of contamination with bacteria due to distance of the well head to septic fields or septic tanks. However, other studies have found an association between the use of a septic system and enteric illness. Denno *et al.* (2009) and Ritter *et al.* (2001) found *E. coli* O157:H7 infections associated with the use of a septic system for home wastewater disposal. In the United States, septic tank and cesspool discharge contribute the highest quantity of wastewater of all sources directly to groundwater, and significant above ground contamination may also occur if leaks or damaged tanks break through the soil surface (Ritter *et al.*, 2002). For this reason, concerns regarding potential risks to human health from this source are predominantly associated with contamination of groundwater with pathogens originating in discharged of leaked septic fluids (Ritter *et al.*, 2002). In a study conducted by Borchardt *et al.* (2003) on septic system density and infectious diarrhoea in children living in Wisconsin, bacterial diarrhoeal illness was found to be associated with holding tank density, with the risk of developing bacterial diarrhoea increasing by 22% for every additional holding tank per quarter-quarter section (40 acre block). However, they failed to find that consumption of well water was the likely transmission route of bacterial infections from nearby septic systems: bacterial pathogens were not isolated from the wells of case households. The authors suggest that contamination of the well water may have been sporadic and that one negative water sample does not necessarily mean a clean water source. Sporadic contamination is even more likely when the size of an aquifer is taken into consideration. The presence of a contaminant in an aquifer near a well may not necessarily be a continuous water quality issue for that well. Large capacity wells that draw in large volumes of clean water along with a contaminant plume will have a large factor of dilution which may decrease the vulnerability of the aquifer when compared to smaller wells which may see little dilution if contaminated (Frind *et al.*, 2006).

In addition to the well type and wastewater management, the household density may also be a risk factor for well water contamination. In developed countries the incidence of enteric illness has steadily declined as a result of improvements in standards of sanitation and hygiene (Rosenberg, 1997). In a study on shigellosis on a First Nations community in Manitoba, the rate for shigellosis was significantly higher where the average household density was 8 or more persons per house (Rosenberg, 1997). These cases are not randomly distributed and instead occur in communities with poor environmental infrastructure (Rosenberg, 1997). In this situation, the control of wastewater and the source of drinking water were likely confounding factors between household density and the incidence of shigellosis.

Geological factors also play a role in the likelihood of a contaminant reaching the ground water source. The soil and subsoil, or vadose zone, is considered a major factor influencing the potential transfer of pathogens through soil to groundwater (Brennan *et al.*, 2010). In a study looking into the ability of *E. coli* to move through temperate maritime soils, the authors found that the greatest number of *E. coli* cells leached from the most poorly drained soils (Brennan *et al.*, 2010). They went on to state that the poorly drained soils had clay content increasing as the depth increased, which likely favoured the formation of macropores. Macropores are known to be important pathways for bacterial transport and reduce the filtration capacity of the soil (Brennan *et al.*, 2010). In another study, Goss *et al.* (1998) found that bacterial contamination in domestic wells in Ontario was greatest in soils that were considered hydrologic group B soils (sandy loams considered moderately permeable). Overall it is believed that the spatial variability in the soil structure may be the most important factor in the transport of bacteria (Brennan *et al.*, 2010).

1.2.2 Land Use Factors

Contaminants that occur in water originate from two primary sources, point source contaminants, such as septic systems as described above, and non-point sources such as diffuse agricultural activity where control of animal wastes generated in different agricultural practices is not contained (Jokinen *et al.*, 2012; Ritter *et al.*, 2002). In the context of pathogenic *E. coli*, cattle are considered to be the primary host for *E. coli* O157:H7 (Karama *et al.*, 2008, Renter *et al.*, 2008; Steinberg *et al.*, 2007; Hussein, 2007, Hancock *et al.*, 2001), but they are not the only reservoir for contamination. Several other species such as rabbits, deer, pigs, chickens and seagulls have also been implicated as carriers of *E. coli* O157:H7 in addition to small domestic ruminants, such as goats and sheep (La Ragione *et al.*, 2009, Hancock *et al.*, 2001). Given the wider availability of sheep and goat products worldwide combined with the intensification of goat and sheep farming, small ruminants could be considered significant reservoirs of *E. coli* O157:H7

(La Ragione *et al.*, 2009). Importantly, all warm blooded animals (i.e., mammals and birds) can act as reservoirs of *E. coli* contamination for groundwater wells used for drinking. Consequently the enteric pathogens associated with animal agricultural sectors as well wildlife represent threats to human health through contamination of drinking water supplies.

Property designation and different measures of livestock density have been found to be an important predictor of risk of *E. coli* contamination of private drinking water (Valcour *et al.*, 2002; Michel *et al.*, 1999). A study conducted by Michel *et al.* (1999) on the geographical distribution of *E. coli* O157:H7 in Ontario found that counties with a very high incidence of verotoxin producing *E. coli* (VTEC) cases, as compared to the Ontario average, were situated in areas of predominately mixed agriculture, where both livestock and the growth of a moderate range of crops took place. In addition, Michel *et al.* (1999) found that the geographic distribution of cattle density presented a similar geographic pattern as the one described for human VTEC cases, with the spatial association between cattle density and VTEC incidence suggesting that living in an agricultural region where cattle are raised could be a risk factor for the acquisition of VTEC disease. Michel *et al.* (1999) went on to state that factors that could be responsible for this association included the contamination of surface water and shallow wells used as drinking water sources and working with or being in close contact with cattle. Their suggestion for this association is not surprising given that *E. coli* O157:H7 can persist and multiply in feedlot soils (Berry *et al.*, 2005). In addition to the presence of livestock, consideration must also be given to the separation between the livestock and the well head for groundwater wells. Goss *et al.* (1998) conducted a study on contamination, defined as the detection of coliforms, fecal coliforms or fecal streptococci in a 250ml sample, in Ontario farmstead domestic wells. The researchers found that on livestock farms, there was a significant decrease in well water contamination by bacteria with increasing separation of the well from the feedlot.

The control of animal wastes generated by agricultural activity has also been identified as a potential non-point source contaminant for drinking water (George *et al.*, 2004, Ritter *et al.*, 2002). A study conducted by Valcour *et al.* (2002) on the associations between indicators of livestock farming intensity and the incidence of human VTEC found that the application of manure to land was identified as a potential risk factor for endemic VTEC infections. Valcour *et al.* (2002) went on to state that runoff from agricultural land that has been treated with manure has the potential to contaminate local surface water and groundwater wells that supply water for human consumption. Outbreaks in both the United Kingdom and Canada have identified well water contaminated with animal manure as the source of illness. Jackson *et al.* (1998) reported on a case of *E. coli* O157:H7 in a 16 month old child living on a cattle farm that was linked to drinking water in the home supplied by a private shallow dug well that was positive for *E. coli*. The characterization of *E. coli* O157:H7 isolates from the well, the case and cattle on the farm were very closely related based on the observation that the pulsed field gel electrophoresis patterns were considered almost identical (Jackson *et al.*, 1998). During the investigation, the farm well head was found to be defective allowing manure-contaminated surface water to flow into the drinking water well (Jackson *et al.*, 1998).

Sewage sludge, or biosolids, is also seen as a potential source of contamination as a result of directly utilizing these products as soil amendments or fertilizers through land application (Horswell *et al.*, 2010; Surampalli *et al.*, 2008). One of the primary concerns with using this resource is the potential transport and survival of pathogenic organisms in soil after sewage sludge application (Horswell *et al.*, 2010). There is a risk of environmental contamination from both horizontal and vertical movement of sewage sludge-borne bacteria (Horswell *et al.*, 2010). Alternatively, a study on the impact of long-term land application of biosolids on groundwater and soil quality at an application site in the United States which had been operated for 8 to 15 years found bacteriological levels in groundwater samples to be close to background levels for fecal coliforms and below permissible limits (Surampalli *et al.*, 2008). The study carried out by Surampalli *et al.* (2008) did not test for any specific pathogenic bacteria, but instead relied on water quality indicator organisms.

1.2.3 Temporal Factors

The role of temporal factors, such as weather, on the incidence of waterborne outbreaks remains a key public health research issue (Curriero *et al.*, 2001). A study conducted by Goss *et al.* (1998) on drinking water wells in Ontario found seasonal variability in the level of contamination, but they also found the subset of wells that were contaminated in the summer was significantly different from the winter subset. The authors suggest that this seasonal variation may have been influenced by soil temperature, manure application schedules and variations in surface hydrology and biological activity related to the growth cycles of the crops. Similarly, a study by Richardson *et al.* (2009) on private water supplies in England found a distinct seasonal effect, with water test results more likely to fail between the months of August and October. Richardson *et al.* (2009) stated that this seasonal effect could be due to changes in land management practices, seasonal agriculture practices, animal movement, with livestock in pastures in the summer but kept indoors in the winter, and climatic factors. Lastly, a study by Schuster *et al.* (2005) on infectious disease outbreaks related to drinking water across Canada also found a distinct seasonal increase with major outbreaks occurring in the spring and summer season. Collectively, these studies demonstrate that bacterial contamination, including contamination with *E. coli*, of drinking water wells is subject to considerable seasonal variability, with groundwater contamination highest during the periods of July to October (Richardson *et al.*, 2009; Schuster *et al.*, 2005; Michel *et al.*, 1999; Goss *et al.*, 1998).

In addition to seasonal variability, a number of climatic factors have been identified as potential contributors to contamination events. A study conducted by Curriero *et al.* (2001) analyzed the relationship between precipitation and waterborne disease outbreaks reported in the United States between 1948 and 1994 and found that 51% of the outbreaks were preceded within a two month time lag by an extreme level of precipitation, defined as precipitation in the 90th percentile for any given location. A similar study by Thomas *et al.* (2006) looked at the role of high impact weather events in waterborne disease outbreaks in Canada between 1975 and 2001 found that rainfall events greater than the 93rd percentile doubled the odds of an outbreak. Schuster *et al.* (2005) also found that extreme weather events were frequently documented contributors in outbreaks related to drinking water in Canada. In addition, Richardson *et al.* (2009) found that the likelihood of well failure increased with increasing intensity of rainfall on the day prior to sampling, with *E. coli* being more likely to be isolated in the water samples if they were obtained one day after a heavy rainfall. The most common threat to the integrity of groundwater sources is contamination resulting from the entry of surface water, either into or down the well casing because of poor construction or maintenance. As such, it is biologically plausible that flooding from extreme precipitation events can increase movement of water and contaminants through the soil profile at a greater than normal rate, resulting in flood waters contaminating groundwater sources (Simpson *et al.*, 2004). It is likely that weather events tend to exacerbate underlying vulnerabilities created by inadequate water protection (Schuster *et al.*, 2005). The outbreak of *E. coli* in Walkerton, Ontario is an example of just such a situation. The presence of an insecure well head, pathogenic bacteria in the environment and inadequate water treatment were exacerbated by heavy rainfall which caused a larger contamination event to occur (Schuster *et al.*, 2005).

1.2.4 Cumulative Impact of Multiple Risk Factors

The relationship between environmental risk factors can also increase the likelihood of groundwater contamination. For example, while the presence of a leaking septic tank is a definite risk factor for the contamination of an adjacent groundwater source, the risk of contamination occurring is further increased during extreme rainfall events and/or when the well head is down gradient and/or close to the source of contamination. The likelihood of groundwater contamination will depend on the intrinsic susceptibility, which is related to the hydrological characteristics of the groundwater source, along with aquifer vulnerability, which in turn is related to the effect of land-use practices and contaminant characteristics and loading (Frind *et al.*, 2006).

While collective studies suggest that bacterial contamination is not uncommon in private water supply systems, efforts to correlate private water supply contamination with predictive factors have been limited (Allevi *et al.*, 2013). To date, little is known about the interactions that multiple environmental risk factors play in facilitating contamination of groundwater. One study carried out on private wells used for drinking water in northeastern Ohio found that the probability of detecting total coliforms was not associated with well depth, age or the township the well was located in (Won *et al.*, 2013). The authors went on to state that the physical factors typically used to identify the characteristics of the wells at high risk for contamination were not predictive of the likelihood of contamination with indicator organisms, *E. coli* O157:H7 or *Campylobacter spp.* (Won *et al.*, 2013) in that geographic area. These authors argued that contamination may be associated with other factors which well owners have greater control over, such as well maintenance. On the other hand, a study looking at private drinking water wells in Pennsylvania found statistically significant associations between the geology, defined as bedrock geology, soil moisture, defined as moisture conditions for the 2 week period prior to sample collection based on the Palmer soil moisture index, and well score, a score created for the presence/absence of 5 important well construction characteristics and the presence of both coliform bacteria and *E. coli* (Swistock *et al.*, 2013). Similarly, a study looking at private wells used for drinking water in Virginia found that 4 variables were significantly associated with total coliform contamination, including the water supply type (ie: well type), the presence of any type of treatment device, the well depth and whether or not the water supply system was located within approximately 1km of a farm animal operation (Allevi *et al.*, 2013). The authors' final regression model included

3 significant predictors of total coliform contamination, well depth, whether there was any type of water treatment device and whether the well was located within 1km of a farm animal operation (Allevi *et al.*, 2013).

As the presence of a fecal indicator bacterium such as *E. coli* suggests an immediate health risk, requiring immediate corrective action, understanding the primary contamination sources would be helpful in the identification of efficient and long-term remediation methods (Allevi *et al.*, 2013). The development of a robust predictive model for groundwater contamination would allow well owners to proactively minimize their risk of well water contamination.

1.3 Objectives and Hypothesis

Private drinking water quality remains a critical public health concern in Canada. Based on the preceding evidence that points to the importance of the interaction between potential environmental risk factors in predicting microbial contamination of drinking water supplies, it is critical that research addresses the cumulative impact that different combinations of environmental risk factors can have on groundwater used as a drinking water source. The objective of this research was to identify significant environmental factors that are associated with the contamination of private drinking water wells with *E. coli*. My general hypothesis was that multiple environmental features interact to affect microbial quality of groundwater used for private drinking water supplies. More specifically, I hypothesized that the following environmental variables played a critical role in the contamination of private drinking water sources with *E. coli* in Southern Ontario and Southern Alberta when compared to uncontaminated private drinking water sources in these same areas:

1) The type of private drinking water well used;
2) the type of wastewater system used;
3) whether livestock was housed on the property within 12 months prior to the collection of the water sample; and,
4) manure storage or spread on the property in the 12 months prior to the water sample being collected.

The work presented in this thesis addresses the risks associated with environmental factors affecting microbial water quality in private drinking water systems in Canada. Chapter 2 describes the methods used throughout the thesis. Chapter 3 presents and discusses the results for environmental risk factors from logistic regression for *E. coli* contamination of private drinking water wells. Chapter 4 presents and discusses the results for environmental risk factors from multiple logistic regression analysis for *E. coli* contamination of private drinking water wells. Lastly, Chapter 5 is a general discussion highlighting the significant findings of this thesis and the importance of these findings to enriching our understanding of the role that the environment plays in contamination of drinking water. This discussion also evaluates the limitations of the current study as well as future research needs.

CHAPTER 2

Research Methods

2.1 Project Overview

The research carried out in this thesis stems from a multi-province surveillance project investigating the prevalence and geospatial distribution of antimicrobial resistant *E. coli*, and funded by the Canadian Institutes for Health Research (CIHR), Health Canada, and the Alberta Heritage Foundation for Medical Research (AHFMR). Between April 2005 and September 2006, a case-control study was conducted using private water samples submitted from participating public health laboratories in Ontario and Alberta. Water surveillance data from the participating public health laboratories were used to identify contaminated wells and the households for a case-control study carried out by Dr. Brenda Coleman. The case-control study, which focused on identifying risk factors for contamination of private water sources with antimicrobial resistant *E. coli*, was carried out in 9 different health regions. Of the nine participating health regions, 7 were in Southern Ontario and included London, Hamilton, Kingston, Orillia, Ottawa and Peterborough and 2 were in Southern Alberta and included Calgary and David Thompson.

The research for this thesis was comprised of the households from the case-control study that reported using a well as their source for drinking water. The focus of this research was to identify environmental variables that play a critical role in the contamination of drinking water wells with *E. coli* in Southern Ontario and Southern Alberta.

2.2 Study Design and Population

The sampling frame for the case-control study consisted of all suitable water samples submitted for bacteriological testing at the participating public health laboratories in Alberta and Ontario between May 1, 2005 and September 30, 2006. A suitable water sample included one submitted from a private drinking water source, collected and sent in an approved bottle that was not broken, leaking or frozen, along with a completed microbial request form for a private drinking water well, and was tested within 24 hours of collection (Technical Advisory Committee on Safe Drinking Water, 2004). As mentioned in Chapter 1, where the water supply is provided as part of the municipal services, the ongoing maintenance and operation are ensured by a public or private utility that must comply with local legislative requirements (Simpson, 2004). When it comes to private water wells, the responsibility for maintenance and operation of the well falls to the well owner. Routine verification of drinking water safety, through the submission of water samples for microbiological testing, is an important component in ensuring the drinking water from the well is safe. The submission of water samples from private wells for testing is voluntary in both Alberta and Ontario. In Ontario, the submission of water samples is done through the local public health unit, with the testing carried out by Public Health Ontario Laboratories. Similarly, in Alberta the submission of water samples is done through Alberta Health Services, with the testing carried out at the Provincial Laboratory of Public Health. While testing recommendations vary, in Alberta, Alberta Health Services recommends that wells that are less than 50 feet deep be tested 4 times a year and wells greater than 50 feet deep be tested 2 times per year.

The case-control study was introduced to potential participants with an information sheet that was mailed along with the bacteriological test results, to all households that submitted a water sample to a participating laboratory during the study period. The cases for the case-control study included households with *E. coli* positive water samples that were resistant to one or more of the antibiotic agents included in the National Antimicrobial Resistance Monitoring System panel for enteric bacteria. The controls consisted of two groups: A) controls which were households randomly selected from the *E. coli* positive samples submitted to the study, that were susceptible to all antibiotics on the screening panel; B) controls which were households that were randomly selected from the provincial database of all water submissions that tested negative for *E. coli* contamination or non-*E.coli* coliform contamination. Both "A" and "B" controls were matched by laboratory region from samples submitted within one month of the data of the case submission. Households selected for inclusion were telephoned by the study assistant. An average of 1.3 controls per case were enlisted to ensure an adequate number of controls were included.

For this thesis, the focus was on *E. coli* contamination, regardless of antimicrobial susceptibility. As such, potential cases were identified as households with *E. coli* contamination of drinking water from their private or communal well (cases and A controls from the base study). Potential controls were identified as households that tested negative for contamination with *E. coli* or non-*E. coli* coliform (B controls from the base study). Both cases and controls were water sources used for drinking.

2.2.1 Inclusion and Exclusion Criteria

Inclusion in this study was limited to households for which the Household Questionnaire was completed as part of the case-control study. Households were not eligible if they did not report having a drilled, dug, bored or driven well. Also ineligible were households that had a water sample that tested positive for *E. coli* contamination within the previous twelve months. Eliminating households that had previously tested positive helps ensure the exposure under investigation occurred prior to the water source becoming contaminated. In doing so, a temporal relationship between the potential exposure and the contamination event was ensured.

2.3 Data Collection

2.3.1 Questionnaires

Households that agreed to participate in the case-control study were asked to provide information about the water source, pets, livestock, and distances between the water source and possible sources of contamination. This information was collected either by telephone interview or in person during a site visit. Each questionnaire represented one household. See appendix 1 for a copy of the questionnaire used.

As data for this study were collected as part of a previous case-control study some of the variables had to be manipulated for use in this research. Where values were missing for the distance between the well head and weeping tile and for the distance between the well head and the septic tank, the median value, stratified by the province was used. For the variables, age of the septic system, depth of the well and age of the well, where responses were missing, the average value, stratified by well type and province was used. For soil type where responses were missing, the soil type was randomly assigned based on the proportion of each soil type reported, stratified by the province. A variable for total income was generated by combining the two categorical variables: 1) reported total off-farm income, and 2) reported total farm income. The mid points of the income ranges reported for the two variables for each household were added together. If either was reported as "don't know" or "not stated", the total income was automatically categorized as "don't know" or "not stated". If either was reported as missing, the total income was reported as missing. Lastly, a categorical variable for the type of septic system used was created. Where more than one type was reported, the one most likely to impact the well water was selected. Details on the variables, including how they were derived, are outlined in Appendix 2.

2.3.2 Meteorological Data

In addition to the information collected from the household questionnaire used in the case-control study, meteorological data from Environment Canada weather stations were obtained based on the county of resident for each household. Five different variables for rainfall data were created based on previous studies completed. A variable for total rainfall on the day the sample was taken, one day prior to the water sample being taken, 2 days prior to the water sample being taken and 3 days prior to the water sample being taken was generated based on the closest weather station for the reported country of residence for each household (Richardson *et al.*, 2009). Unlike the study conducted by Richardson *et al.* (2009) where rainfall amounts ranged between less than 1mm to over 1000mm, during the sampling period for this research, low amounts of rainfall were recorded. The mean daily totals ranging between 2mm and 3mm with the majority of days having no rainfall reported. As a result, the variable for total rainfall on each of the days was converted into a dichotomous categorical variable looking at rainfall versus no rainfall on the day. The last variable created was a categorical variable that summed the total rainfall for the 3 days prior and the day of the water sample collection.

2.3.3 Animal Density

An animal density calculation was completed for each household as an indicator of the pressure of livestock farming on the surrounding environment. Animal density was calculated as the number of animal units per square kilometre. A categorical variable identifying animal density for each household as low, medium and high, based on established cut points, was generated (Minnesota Department of Agriculture, 2005). The calculation of animal units was based upon a published measure widely used to estimate the feed inputs and manure outputs of farm animals (Minnesota Department of Agriculture, 2005). The area for the property for each household was taken from the Household Questionnaire.

2.4 Data Analysis

The dependent variable represented a household's laboratory test result for their drinking water and designated as a binary value: *E. coli* absent (0) or present *E. coli* present (1) in the well water sample they submitted for analysis.

The independent variables of primary interest were:

1. The type of well used:
    1) Drilled

2) Bored

3) Driven

2. The type of septic system used:

    1) Septic tank and weeping bed

    2) Field tank

    3) Holding tank

    4) Lagoon

    5) Surface discharged

    6) Municipal

    7) Other

3. Livestock housed on the property in the past 12 months

    1) Yes

    2) No

4. Manure storage or spread in the past 12 months

    1) Within 15m of the well

    2) Within 30m of the well

    3) Within 100m of the well

    4) More than 100m from the well

    5) Not spread in the past 12 months

    6) Don't know or refused.

2.4.1 Variable Identification:

The goal of this analysis was to find the best and most biologically credible way to describe the relationship between the outcome (i.e. presence or absence of *E. coli* in the well water sample) and a set of risk factors or independent variables, which included the type of septic system, water well and exposure to animal contamination (Figure 1). The independent variables were selected based on a review of the current literature.

Effect modifiers are variables that are identified as interacting with a risk factor (Hosmer *et al.*, 2000). When the association between an independent variable and the dependent variable differs in some way depending on the level of a third variable, interaction is present (Hosmer *et al.*, 2000). The third variable is called the effect modifier (Aschengrau and Seage, 2008). The potential effect modifiers, identified through the literature search, included: 1) interaction between manure storage in the past 12 months and livestock housed on the property in the past 12 months; 2) interaction between livestock housed on the property in the past 12 months and the type of well used; 3) interaction between the age of the well and the depth of the well; and 4) interaction between total rainfall over the four days prior to the water sample collection and the season the sample was taken in.

Hosmer *et al.*. (2000) describes a confounder as a covariate that is associated with both the outcome variable of interest and primary independent variables of interest. Confounding is the mixing of effects between an exposure, and outcome, and a third extraneous variable known as the confounder (Aschengrau and Seage, 2008). When confounding is present, the association between the exposure and the outcome is distorted because of the relationships between the confounder and the exposure, and between the confounder and the outcome or disease (Aschengrau and Seage, 2008). If confounders are not taken into account when analysing the association between the independent variables and the dependent variables, the distribution of the confounder can artificially inflate or deflate the main association under investigation. There were five potential confounders identified for this study. Geological factors play a role in the likelihood of a contaminant reaching the ground water source. Goss *et al.* (1998) found that bacterial contamination in domestic wells in Ontario was greatest in soils that were sandy loams and considered moderately permeable. The more permeable the soil surrounding the septic system and the drinking water well, the easier it is for contaminants to move through the soil profile. In addition, the type of soil may also influence the type of well installed. The presence of sandy soils and a high water table may result in the installation of a dug or driven well, while bedrock and a deep water table will require the installation of a drilled well. As a result the soil type where the drinking water well is located may act as a potential confounder for this study. Secondly, contaminants that occur in water originate from two primary sources, point source contaminants, such as septic systems, and non-point sources such as land use practices including agricultural activity and the control of animal wastes generated in different agricultural practices (Ritter *et al.*, 2002). Wells located in areas that are associated with a higher number of potential routes for contamination are more likely to be contaminated, making property designation a potential confounder. In addition, total income and education may play a role in the understanding and ability to maintain drinking water wells. Well owners are principally responsible for maintaining and preventing contamination of their wells (Kreutzwiser *et al.*, 2011). In turn, the responsibility to effect source water protection through proper management of potential contaminants falls to the well owners themselves (Kreutzwiser *et al.*, 2011). This requirement for protection necessitates an understanding of how potential sources of contamination may impact a drinking water well, as well as the financial capability to properly maintain the well. As such, total income reported for the household and highest level of education attained may represent potential confounders, and were evaluated as such in this study. Lastly, province (Ontario or Alberta) was also identified as a potential confounder. Households from two different provinces were used in this study. Ontario and Alberta differ with respect to land uses, including the type and amount of agricultural activity that takes place within each province. In addition, the method by which well water samples are analyzed for the presence of *E. coli* differs between Ontario and Alberta. These differences remain unmeasured and may impact the identification of *E. coli* in drinking water samples submitted to the respective public health laboratories in the two provinces.

In addition to effect modification and confounding, potential sources of bias were also considered. While bias can occur in all types of epidemiologic studies, retrospective studies are more susceptible to bias than prospective ones (Aschengrau and Seage, 2008). The two main types of bias that can occur include selection bias, which occurs during the selection and follow-up of participants in the study, and observational bias, which occur during data collection (Aschengrau and Seage, 2008). A survey of over 1000 private well owners in Ontario found that while almost all respondents reported having has their private well water tested at least once (94%), 65% reported testing their water less than once a year (Kreutzwiser *et al.*, 2011). Those findings were similar to smaller studies completed in Ontario which found between 61% of private well owners did not test their water annually (Kreutzwiser *et al*., 2010) and 75% of private well owners did not test their water annually (Hexemer *et al.*, 2008). Kreutzwiser *et al.*, (2011) found that respondents who were less concerned about water safety seemed less inclined to submit a water sample for testing. As households for this research were identified through the voluntary submission of a water sample to a participating public health laboratory, participants willing to be included in this study may have differed from those that chose not to participate, introducing volunteer bias into the study. In order to limit volunteer bias, recruitment for the study covered wide geographical areas of both Ontario and Alberta, and included farming and non-farming properties.

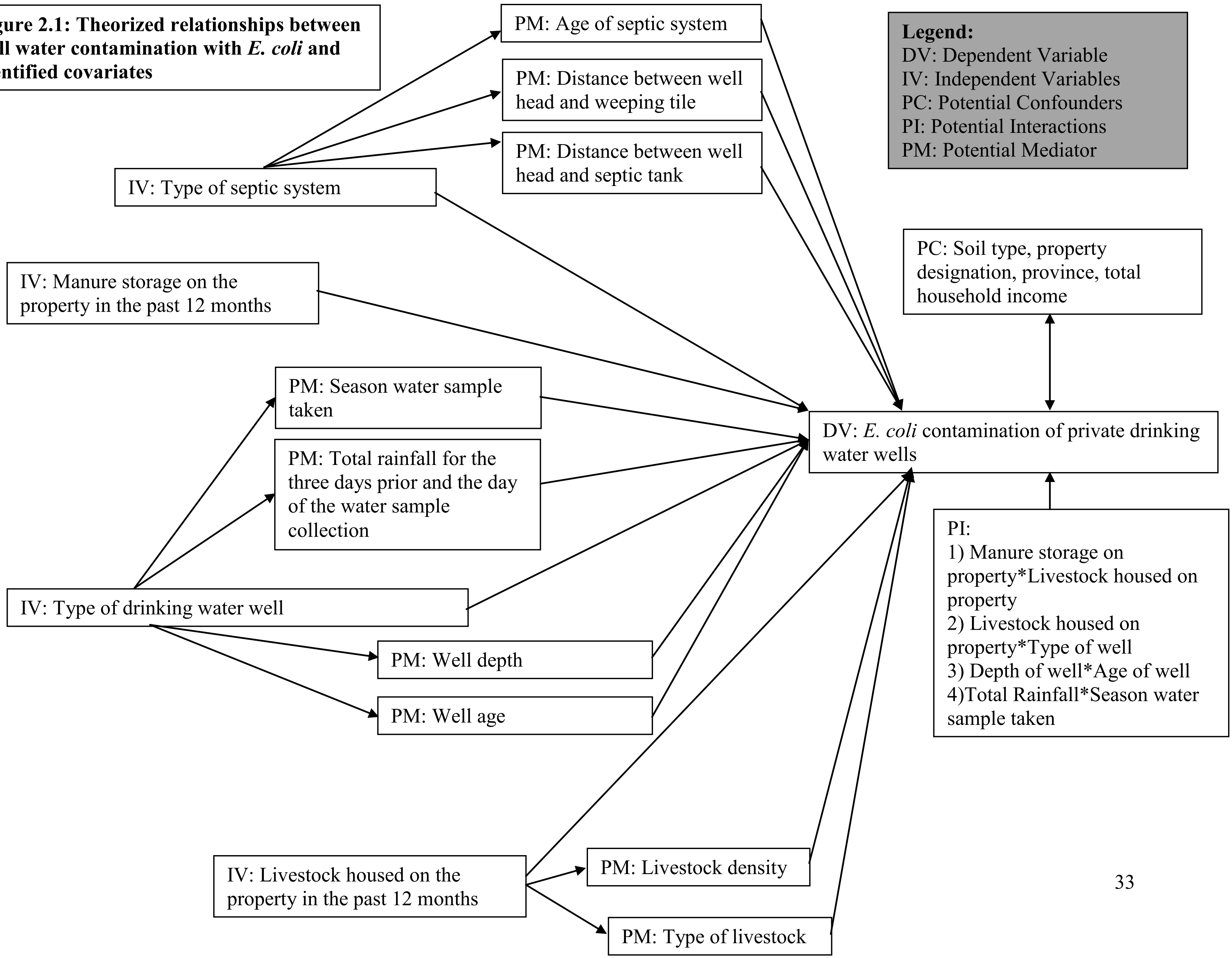
In addition, information about the mode of interview (site visit or telephone interview) was included in the analysis. It has been previously demonstrated that different modes of questionnaire administration are likely to affect the quality of data collected (Bowling, 2005; Catania *et al.*, 2012). There can be highly situational effects of interviewing, which may depend on complex interplay between modes of interview, person variables and the interview topic itself (Catania *et al.*, 2012). As such, the mode of interview was retained as a covariate for analysis.

Lastly, mediation analysis was undertaken to help understand how the independent variables affected the outcome. A mediator is defined as the mechanism through which an independent variable can influence the dependent variable (Fraszier *et al.*, 2004; Baron *et al.*, 1986). Statistically mediation and confounding are very similar, but the conceptualizations underlying the mediational and confounding hypothesis are quite different (MacKinnon *et al.*, 2000). Mediation involves a distinctly causal relationship among the variables, and the direction of causation involved in the mediated effect implies that the independent variable causes the mediator, which in turn causes the dependent variable (MacKinnon *et al.*, 2000). The confounding hypothesis, on the other hand, focuses on adjustment of observed effects to examine undistorted estimates of effect (MacKinnon *et al.*, 2000). The distinctions between mediation and confounding involve the directionality and causal nature of the relationships in the model, and these particular aspects of model specification are not determined by statistical testing (MacKinnon *et al.*, 2000). In examining mediation, the relationship between an independent variable and a dependent variable is broken down into two causal pathways, one that links the independent variable to the outcome variable directly, and the other that links the independent variable to the outcome variable through a mediator (MacKinnon *et al.*, 2000).

For this thesis, contamination of well water with *E. coli* may be associated with the type of well, in turn this relationship may be mediated by the age and depth of the well, rainfall on the day the water sample was collected, the total rainfall over the three days prior to and the day of the water sample collection, and/or the season the water sample was taken. In addition, the relationship between well water contamination with *E. coli* and the type of septic system on a property may be mediated by the distance between the well head and the septic tank, the distance between the well head and weeping tile, and/or the age of the septic system. Lastly, contamination of well water with *E. coli* may be associated with housing livestock on the property. This relationship may be mediated by the type of livestock housed and/or the livestock density. These potential mediational relationships were considered in this study.

The theorized relationship between well water contamination with *E. coli*, the type of septic system used, the type of well used, manure storage in the past 12 months, livestock housed on the property in the past 12 months, and identified effect modifiers, confounders and mediators are presented below in Figure One.

Figure 2.1: Theorized relationships between well water contamination with *E. coli* and identified covariates



2.4.2 Model Building Strategies:

The development of an appropriate regression model depends on the research objective identified. Model building can fall into one of two categories, explanatory or exploratory (Kleinbaum *et al.*, 2008). Explanatory model building starts with a hypothesized association that is tested to determine whether the association remains or the strength is affected when other factors (identified confounders, moderators, and/or effect modifiers) are controlled for (Courvoisier *et al.*, 2011; Williams *et al.*, 2001).

Alternatively, the goal of exploratory model building is not to examine theoretically specified relationships, but instead to find a good set of predictors of an outcome. When using an exploratory model building strategy the emphasis is on identifying which covariates should be included in the model that will best describe the relationship between the outcome and a set of independent variables (Courvoisier *et al.*, 2011; Williams *et al.*, 2001).

The goal of this analysis was exploratory: to identify the relationship between the outcome, which was the presence or absence of *E. coli* in the well water sample, and potential risk factors, including well characteristics and exposure to human or animal contamination, as outlined in Figure 1. The strategy outlined by Hosmer *et. al*. (2000) was used and all variables that were statistically significant and those which were identified as biologically important were included. Since the dependent variable was dichotomous (presence/absence of *E. coli*) and the study design was case-control, logistic regression was used to analyze the data. To properly quantify the relationship between the identified independent variables of interest and the dependant variable, effect modifiers, potential confounders and mediators were also considered.

Variables were selected based on associations identified using logistic regression. The Chi-square test was used for categorical variables and the two-sample t-test was used for normally distributed continuous variables. Continuous variables that were not normally distributed were converted into categorical variables. Categories were assigned based on common cut-offs identified in the literature and/or legislation. Continuous variables converted into categorical variables were analyzed using the Chi-square test.

After completing the bivariate analysis, variables were selected for multivariable model building. Variables were included in the full model if it had a p-value of ≤ 0.25 from the bivariate test, if it was related to the study design or if it was considered biologically important. As recommended by Hosmer *et al.* (2000), a p-value of 0.25 was chosen because a more traditional level, such as 0.10, may fail to identify variables known to be important. A multivariable model with these variables and the design-related was then fit.

Prior to building the final model, assessment for collinearity was carried out using the Cramer's V statistic, which is a measure of association between two nominal variables. The statistic goes from 0 to 1, with 1 indicating a strong association, and the likelihood that the 2 variables are measuring the same concept. This statistic was chosen as it can be used with categorical variables. Using this statistic, a number of variables were identified as collinear. In these cases, the most biologically relevant variable was included, the other collinear variable dropped and the reduced multivariable model run again. Variables from the multivariable model with a p-value ≤ 0.05, and those identified as biologically significant were retained for the main effects model.

The confounding effects of the variables removed were assessed by determining whether the change in the regression coefficients of the primary predictors, which stayed in the reduced model, changed by more than 15% when the variables were removed (Hosmer *et al.*, 2000). If any of the regression coefficients changed by more than 15%, the removed variable was included back in the model. The new model was compared to the older, larger model using the likelihood ratio test. This process was continued until all the clinically and statistically significant variables were included in the model (Hosmer *et al.*, 2000).

Analysis of the mediation effects of potential mediators identified earlier was also assessed. The mediation effect of variables identified as potential mediators that remained in the model following Bivariate analysis and assessment for confounding was carried out using the strategy outlined by Baron *et al.* (1986). Baron *et al.* (1986) recommended running three regression equations for each potential mediator:

1) Regression of the mediator on the independent variable,
2) Regression of the dependent variable on the independent variable, and
3) Regression of the dependent variable on both the independent variable and on the mediator.

To establish mediation, Baron *et al.* (1986) states the following conditions must hold:

1) The independent variable must affect the mediator in the first equation, meaning there must be a significant relationship between the independent variable and the dependent variable.

2) The independent variable must be shown to affect the dependent variable in the second equation, meaning there must be a significant relationship between the independent variable and the mediating variable.
3) The mediator must affect the dependent variable in the third equation, meaning the mediator must be a significant predictor of the dependent variable in an equation including both the mediator and the independent variable.

If these conditions all hold, then the effect of the independent variable on the dependent variable must be less in the third equation than in the second, in other words, if the mediator were to be removed, the relationship between the independent and dependent variables would be significantly reduced. Variables identified as mediators were retained in the main effects model.

The linear assumption of continuous variables was then assessed using the method outlined by Hosmer *et al.* (2000). Hosmer *et al.* (2000) states that assuming linearity at the variable selection stage is consistent with the goal of determining whether a particular variable should be in the model. Once it is determined the variable should be included in the model, the linear assumption can be assessed. Descriptive statistics were used to obtain the quartiles of the distribution of the continuous variables. A categorical variable was then created with 4 levels using the three cut points based on the quartiles. A multivariable model replacing the continuous variable with the four-level categorical variable was fit. Following the fit of the model, the estimated coefficients versus the midpoints of the groups was plotted. The plot was visually inspected to determine if a linear relationship existed. If the relationship was not linear, the continuous variable was grouped into categories based on quartiles and fitted as a nominal variable. The remaining variables, identified confounders and mediators made up the main effects model.

Once the main effects model was complete, biologically plausible interactions among variables in the main effects model were assessed. Interaction variables are created as the arithmetic product of the pairs of main effect variables (Hosmer *et al.*, 2000). The potential interaction variables were added into the model one at a time. Any interaction variables that did not remain statistically significant at $p \leq 0.05$ when included in the model were dropped.

The remaining variables, identified confounders, mediators and interaction variables were then fit into the final model. The final model was assessed to test the goodness of fit using the Hosmer-Lemeshow tests (Hosmer *et al.*, 2000).

CHAPTER THREE

An Evaluation of Environmental Factors Affecting Microbial Contamination of Drinking Water Wells Using Bivariate Analysis

3.1 Introduction

Investigations to identify the routes of drinking water contamination are necessary to better understand the role of microbial transmission from the environment and to help focus the implementation of evolving preventative measures. Environmental risk factors for the contamination of groundwater used as a private drinking water source with *E. coli* can be divided into three main categories:

1) physical factors,
2) land use factors, and;
3) temporal factors.

The objective of the research presented in this chapter was to present the results of the bivariate analysis, identifying significant environmental factors that were associated with the contamination of private drinking water wells with *E. coli*. Risk factors identified as significant in this chapter were then used to form the multivariable model described in Chapter 4.

3.2 Results and Discussion

3.2.1 Characteristics of Households

In total, 1,157 households were administered the Household Questionnaire during the study period. Potential households for inclusion in the study were identified by water samples that were tested for bacterial contamination by participating public health laboratories. Of the original 1,157 households enrolled in the case-control study, 129 households were excluded from the study because of a previous positive water sample for *E. coli* or coliform bacteria in the 12 months prior to the interview being conducted. In addition, 173 households were excluded because they did not report using a private well (i.e., used shore/lake well, cistern, spring, or did not know) when being interviewed. The final analysis for this research was conducted using the household questionnaires collected from 855 households, each representing one well and property. Of the 855 households, there were 534 cases and 321 controls with a ratio of approximately 1.7 cases for every control.

Table 3. 1: Households Included, Ontario & Alberta, 2005-2006

<table>
<tr><td colspan="2" rowspan="2">Water sample testing for bacterial contamination by participating public health laboratories and enrolled in case-control study</td><td colspan="2">Households Included</td></tr>
<tr><td>Cases</td><td>Controls</td></tr>
<tr><td>Household with E. coli positive water sample</td><td>Household with bacteria free-water sample</td><td rowspan="4">680</td><td rowspan="4">477</td></tr>
<tr><td>⇩</td><td>⇩</td></tr>
<tr><td>Potential Case</td><td>Potential Control</td></tr>
<tr><td>⇩</td><td>⇩</td></tr>
<tr><td colspan="2">Removed households that had a water sample test positive for bacteria in the 12 months prior.
Case households=23 Control households=106</td><td rowspan="2">657</td><td rowspan="2">371</td></tr>
<tr><td colspan="2">⇩</td></tr>
<tr><td colspan="2">Removed households that did not report using a private drinking water well as their water source.
Case households=123 Control households=50

Answer to Q18: “What type of well do you have? Is it drilled, dug, bored or driven, which is also called a sandpoint or well point?” = Other, Don’t know, or was not answered by the respondent.</td><td>534</td><td>321</td></tr>
</table>

Table 3.2 highlights the demographic information about the 855 households included in this study. Five hundred thirty four (63%) of the households had private well

water contaminated with *E. coli*, and three hundred twenty one (38%) of the households did not have any contamination. This grouping is a product of the selection of households for the case-control study completed as part of the original research project and is not the probability of contamination of the private drinking water wells. Household sizes ranged from one to sixteen individuals with a median number of two people per household (mean=2.8).

Of the responders from each household, 55% were male and 45 % were female. Responders ranged in age between 21 years and 93 years, with a mean age of 54 years (median=55 years). There was no difference between cases and controls with respect to the responder's age or gender. The majority of responders (75%) reported living on the property for 5 or more years, with a mean of 15 years (median= 11 years). There was no difference between cases and controls with respect to the amount of time the household had resided on the property.

Of the households that were included in the study, 415 (49%) had one or more household members with a college or university degree while 465 households (54%) reported a total income of $40,000 or greater, although 286 households (33%) did not state their total income. There was no significant difference between contaminated and uncontaminated wells and the distribution of education of the household members or household income.

Two hundred and forty three (28%) of the households were located in the province of Alberta and 612 (72%) of the households were located in the province of Ontario. A higher proportion of *E. coli*-contaminated wells were located in the province of Ontario (74%) versus Alberta (26%) when compared to wells that were not contaminated with *E. coli*, and had not been for one year prior to the well water sample being collected. Lastly, 130 (15%) of the households were interviewed for this study via an on-site visit and 725 (85%) of the households were interviewed via phone. There was no significant difference between contaminated and uncontaminated wells and the mode of interview.

Table 3. 2: Demographic information for the study households using private drinking water wells from Southern Ontario & Southern Alberta (2005-2006)

Variable	**All Households N=855 (% Households)**	**Cases N=534 (% Cases)**	**Controls N=321 (% Controls)**	**p-value**
Number of Household Members				
Mean/ Median	2.8/2	2.9/2	2.8/2	0.414
Not stated	1 (0.1)	1 (0.2)	0 (0)	
Highest Level of Education in Household				
Less than high school graduation	60 (7.0)	43 (8.1)	17 (5.3)	0.399
Graduated high school	144 (16.8)	85 (15.9)	59 (18.4)	
College or trade school	271 (31.7)	171 (32.0)	100 (31.2)	
University	346 (40.5)	215 (40.3)	131 (40.8)	
Not stated	34 (4.0)	20 (3.8)	14 (4.4)	
Total Household Income				
Less than $40 000	104 (12.2)	76 (14.2)	28 (8.7)	0.124
$40 000 to less than $60 000	93 (10.9)	57 (10.7)	36 (11.2)	
$60 000 to less than $80 000	100 (11.7)	65 (12.2)	35 (10.9)	
$80 000 or more	272 (31.8)	159 (29.8)	113 (35.2)	
Don't know, refused or not stated	286 (33.5)	177 (33.2)	109 (34.0)	
Province				
Alberta	243 (28.4)	139 (26.0)	104 (32.4)	0.046
Ontario	612 (71.6)	395 (74.0)	217 (67.6)	
Interview Mode				
Telephone	725 (84.8)	450 (84.3)	275 (85.7)	0.581
Site visit	130 (15.2)	84 (15.7)	46 (14.3)	

The odds ratios for dichotomous and categorical variables identified as potential confounders or covariate included to assess for sources of bias are presented in Table 3.3. There was no significant association between the household's total reported income, education or the mode of interview and *E. coli* contamination of the water well. The province in which the household was located was associated with water contamination with *E. coli*. Participating households in Ontario were more likely to have an adverse water result as compared to households in Alberta (OR, 1.36 [95% CI, 1.01-1.84]). This difference between provinces with respect to *E. coli* contamination could have been due to the differences between the methods used for testing in the respective provincial laboratories.

Table 3. 3: Bivariate associations between E. coli contamination of ground water used as a private drinking water source and covariates included to assess for bias and potential confounding in the survey of households using private drinking water wells in Southern Ontario & Southern Alberta (2005-2006)

Variable	Odds Ratio	95% CI	p-value
Total Household Income (ref: $80 000 or more)			
Less than $20 000 to less than $40 000	1.93	1.17-3.17	0.009
$40 000 to less than $60 000	1.13	0.69-1.82	0.631
$60 000 to less than $80 000	1.32	0.82-2.13	0.254
Don't know or refused or not stated	1.15	0.82-1.62	0.408
	Overall		0.114
Highest Education in Household (ref: Graduated high school)			
Less than grade 9 or Some high school	1.76	0.91-3.37	0.091
College or trade school	1.19	0.78-1.80	0.417
University	1.14	0.76-1.69	0.520
	Overall		0.387
Province (ref: Alberta)			
Ontario	1.36	1.01-1.84	0.046
Mode of interview (ref: Telephone)			
Site	1.12	0.76-1.65	0.581

3.2.2 Physical Risk Factors – Results and Discussion

Table 3.4 outlines the different indicators of physical risk factors analyzed for association with *E. coli* contamination of the well water. Of the 855 households, the majority (73%, N=622) reported using a drilled well, 24% (N=211) reporting using a dug or bored well and 3% (N=22) reported using a driven well. A higher proportion of *E. coli*-contaminated wells were dug or bored wells (31%) compared with wells that were not contaminated and had not been for one year prior to the well water sample being collected (15%). Well depth ranged between <1m to 203m, with a median depth of 23m (mean=28m). The majority of *E. coli*-contaminated wells were shallower than 30m (65%) compared to wells that were not contaminated (48%).

Well age ranged between <1 year to 160 years, with a median age of 22 years (mean=26 years). The majority of *E. coli*-contaminated wells were older than 25 years (53%) compared to wells that were not contaminated (37%).

The wastewater management system used by the majority of the households was a septic tank and weeping bed (93%, n=794). Of the remaining households, field tanks (n=5), holding tanks (n=12), lagoons (n=3), surface discharge (n=13) and municipal waste water management (n=14) were reported as the primary waste water management system used, while 9 households reported using something other than one of the options listed as their wastewater management system and 5 households reported not knowing what method they used for wastewater management. There was no difference between contaminated and uncontaminated wells with respect to the type of wastewater management system used on the property (p-value=0.867). The age of wastewater management systems used on the property ranged between <1 a year to 100 years, with a median age of 19 years (mean=20years). The average age of the waste water management systems for households with *E. coli*-contaminated water wells was older (mean age=21.3 years) when compared with households that did not have *E. coli*-contaminated water wells (mean age=19.0 years; p-value=0.017). The distance between the well head and the septic tank ranged between <1m to 1609m, with the median distance of 30m (mean=58m). Similarly, the distance between the well head and the weeping tile ranged between 2m to 1600m, with a median distance of 31m (mean=61m). There was no difference between contaminated and uncontaminated wells and the distance between the well head and the weeping tile (p-value=0.788) nor the distance between the well head and the septic tank on the property (p-value=0.509).

Of the 855 households, 36% reported clay as the predominant soil type on their property, 28% reported loam, 17% reported sand and 16% reported gravel. There was no difference between contaminated and uncontaminated wells and the reported predominant soil type for the property (p-value=0.078).

Table 3. 4: Comparison of identified physical risk factors for the study households using private drinking water wells from Southern Ontario & Southern Alberta (2005-2006)

Variable	All Households N=855 (% Households)	Cases N=534 (% Cases)	Controls N=321 (% Controls)	p-value
Depth of well (m)				
Min to 10	208 (24.3)	148 (27.7)	60 (18.7)	
10.1 to 30	298 (34.9)	201 (37.6)	97 (30.2)	
30.1 to 40	156 (18.3)	86 (16.1)	70 (21.8)	
40.1 to max	193 (22.6)	99 (18.5)	94 (29.3)	<0.001
Type of Well				
Drilled	622 (72.8)	364 (68.2)	258 (80.4)	
Dug or bored	211 (24.7)	163 (30.5)	48 (15.0)	
Driven	22 (2.6	7 (1.3)	15 (4.7)	<0.001
Age of well (years)				
Min to 10	195 (22.8)	103 (19.3)	92 (28.7)	
10.1 to 25	255 (29.8)	146 (27.3)	109 (34.0)	
25.1 to 35	236 (27.6)	165 (30.9)	71 (22.1)	
35.1 to max	169 (19.8)	120 (22.5)	49 (15.3)	<0.001
Distance between well head and weeping tile (m)				
Min to 20	153 (17.9)	94 (17.6)	59 (18.4)	
20.1 to 35	296 (34.6)	190 (35.6)	106 (33.0)	
35.1 to 60	117 (13.7)	69 (12.9)	48 (15.0)	
60.1 to max	289 (33.8)	181 (33.9)	108 (33.6)	0.788
Distance between well head and septic tank (m)				
Min to 20	255 (29.8)	163 (30.5)	92 (28.7)	
20.1 to 30	163 (19.1)	95 (17.8)	68 (21.2)	
30.1 to 50	224 (26.2)	137 (25.7)	87 (27.1)	
50.1 to max	213 (24.9)	139 (26.0)	74 (23.1)	0.509

Table 3.4 continued...

Variable	All Households N=855 (% Households)	Cases N=534 (% Cases)	Controls N=321 (% Controls)	p-value
Type of Septic System				
Septic tank and weeping bed	794 (92.9)	496 (92.9)	298 (92.8)	
Field tank	5 (0.6)	4 (0.8)	1 (0.3)	
Holding tank	12 (1.4)	7 (1.3)	5 (1.6)	
Lagoon	3 (0.4)	1 (0.2)	2 (0.6)	
Surface discharge	13 (1.5)	9 (1.7)	4 (1.3)	
Municipal	14 (1.6)	8 (1.5)	6 (1.9)	
Other	9 (1.1)	5 (0.9)	4 (1.3)	
Missing	5 (0.6)	4 (0.8)	1 (0.3)	0.867
Age of septic system (mean years)	20.46	21.27	19.10	0.017
Soil type				
Gravel	145 (17.0)	94 (17.6)	51(15.9)	
Sand	153 (17.9)	86 (16.1)	67 (20.9)	
Loam	242 (28.3)	143 (26.8)	99 (30.8)	
Clay	315 (36.8)	211 (39.5)	104 (32.4)	0.078

The odds ratios for dichotomous and categorical variables identified as physical risk factors investigated are presented in Table 3.5. For type of well used, there was a significant difference between the reference group, households that used a drilled well, and those which used a dug well (OR 2.41 [95% CI,1.68-3.45]), or a sandpoint well (OR 0.33 [95% CI, 0.13-0.82]). The odds of a water sample submitted from a dug well being contaminated with *E. coli* was two times higher when compared to drilled wells. On the other hand, the odds of a water sample submitted from a driven or sandpoint well being contaminated with *E. coli* was only 1/3 when compared to drilled wells.

For well depth, there was a significant difference between the reference group, households with a well depth greater than 40.1m, and households with a well depth of ten meters or less (OR 2.34 [95% CI, 1.55-3.54]) as well as households with a well depth between 10.1m and 30m (OR 1.97 [95% CI, 1.36-2.85]). The odds of a water sample submitted from a well that was less than or equal to 30m deep being contaminated with *E. coli* was approximately two times higher when compared to wells that were greater than forty meters deep. There was no difference between the odds of *E. coli* contamination for households with wells that were between 30.1m and 40m and households with wells over 40m deep.

Well depth is very closely tied to the type of well. As mentioned in Chapter 1, section 1.2, dug wells are shallow holes that are dug approximately 3 to 9 meters deep, driven wells are approximately 15 meters deep and drilled wells are typically between 30 to 120 meters deep (Committee on Environmental Heath and Committee on Infectious Disease, 2009). Given this, it was expected that the odds of *E. coli* contamination would be higher for households that reported using a dug well and for well depths reported to be less than 30m deep, as was found in this sample. These findings are in line with what Goss *et al.* (1998) reported. In their research, they determined that the combination of well type and depth accounted for the largest reduction in contamination, with dug or bored wells being more frequently associated with contamination than drilled wells, which are typically more than 30 meters deep. Similarly, a study carried out by Audette *et al.* (2001) on farm well water quality in Alberta found that shallow wells, which were classified as less than 30m deep, were more likely to be contaminated with coliform bacteria than wells 30m or deeper.

For the age of the well, there was no significant difference between the reference group, well age less than ten years, and households with a well between ten and twenty five years of age. The odds of *E. coli* contamination was significant for households with a well that was between 26 years to 35 years old (OR 2.07 [95% CI, 1.40-3.08]) and wells that were greater than 36 years of age (OR 2.18 [95% CI, 1.42-3.38]). The odds of a water sample submitted from a well that was twenty six years or older being contaminated with *E. coli* was two times higher when compared to wells that were less than ten years. This finding is similar to what was reported by Goss *et al.* (1998), who found a statistically significant increase in the number of bacterial colonies as the age of the well increased.

The type of septic system used was not significantly associated with *E. coli* contamination of the well water in our sample. Previous research has shown that the risk of ground water contamination from private wastewater systems that consist of surface discharge or pits and lagoons is significantly higher than the use of a septic tank with a leaching field (Shook *et al.*, 1993). The likelihood that a septic tank with a leaching field will become a risk factor in groundwater contamination is related more to the installation of the system and ensuring proper maintenance versus the presence of the system in the first place (Mathis *et al.*, 2011). While the type and location of the septic system used was not associated with *E. coli* contamination, the age of the septic system was. For every ten year increase in the age of the septic system, the odds of *E. coli* contamination increased by a factor of 13% (OR, 1.13 [95% CI 0.032-1.25]). This finding, that age of the system rather than the type of the system, is associated with *E. coli* contamination may be explained given the older the system, the more likely it is to fail and thus become a potential point source of pollution. The ongoing maintenance and operation of the septic system is the responsibility of the septic system owner. The result is that many private septic systems may not be properly maintained while in use and are thus at an increased risk for failure. Malfunctioning or poorly maintained septic systems have been the cause of outbreaks of waterborne disease in the United States attributed to contaminated ground water (Ritter *et al.*, 2002; Borchardt *et al.*, 2003). In addition to the type of wastewater collection system used, neither the distance between the well head and the septic tank (p-value = 0.511) nor the distance between the well head and the weeping bed (p-value = 0.789) were significantly associated with *E. coli* contamination of the well water. Similarly, Goss *et al.* (1998) found no significant effects on the level of contamination with bacteria due to distance of the well head to septic fields or septic tanks. In addition, Swistock *et al.* (2013) found no association between *E. coli* contamination of water wells and the estimated distance between the water well and the septic system.

For soil type, there was no significant difference between the reference group, households that reported sand and households that reported gravel or loam. The odds of *E. coli* contamination was significant for households that reported clay (OR 1.58 [95% CI 1.06-2.35]). This association with clay is similar to results from previous research. A study investigating the ability of *E. coli* to move through temperate maritime soils found that the greatest number of *E. coli* cells leached from the most poorly drained soils. (Brennan *et al.*, 2010) The authors went on to state that the poorly drained soils had clay content increasing as the depth increased, which likely favored the formation of macropores. Macropores are known to be important pathways for bacterial transport and reduce the filtration capacity of the soil (Brennan *et al.*, 2010).

Table 3. 5: Bivariate associations between E. coli contamination of ground water used as a private drinking water source and covariates identified as physical risk factors in the survey of households using private drinking water wells in Southern Ontario & Southern Alberta (2005-2006)

Variable	Odds Ratio	95% CI	p-value
Well depth (m) (ref: 40.1 to max)			
min to 10	2.34	1.55-3.54	<0.001
10.1 to 30	1.97	1.36-2.85	<0.001
30.1 to 40	1.17	0.76-1.78	0.476
	Overall test	-	<0.001
Well type (ref: Drilled wells)			
Dug or Bored	2.41	1.68-3.45	<0.001
Sandpoint	0.33	0.13-0.82	0.017
	Overall test	-	<0.001
Age of Well (years) (ref: min to 10 years)			
10.1 to 25	1.19	0.82-1.74	0.349
25.1 to 35	2.08	1.39-3.08	<0.001
35.1 to max	2.19	1.42-3.38	<0.001
	Overall test	-	<0.001
Distance between well head and weeping tile (ref: 60.1 to max)			
min to 20	0.95	0.63-1.42	0.806
20.1 to 35	1.07	0.76-1.49	0.695
35.1 to 60	0.86	0.55-1.33	0.493
	Overall test	-	0.789

Table 3.5 continued...

Variable	Odds Ratio	95% CI	p-value
Distance between well head and septic tank (ref: 50.1 to max)			
min to 20	0.94	0.64-1.38	0.763
20.1 to 30	0.74	0.49-1.13	0.167
30.1 to 50	0.84	0.57-1.23	0.375
	Overall test	-	0.511
Type of septic system (ref: Septic tank and weeping bed)			
Field tank	2.40	0.27-21.60	0.434
Holding tank	0.84	0.26-2.67	0.769
Lagoon	0.30	0.02-3.33	0.327
Surface discharge	1.35	0.41-4.43	0.619
Municipal	0.80	0.28-2.33	0.684
Other	0.75	0.20-2.81	0.671
	Overall test	-	0.874
Age of Septic System (10 years)	1.13	0.032-1.25	0.018
Soil Type (ref: sand)			
Gravel	1.44	0.90-2.29	0.129
Loam	1.13	0.75-1.69	0.572
Clay	1.58	1.06-2.35	0.024
	Overall test	-	0.078

3.2.3 Land Use Risk Factors – Results and Discussion

Table 3.6 outlines the different land use risk factors that were analyzed for association with *E. coli* contamination of the well water. Of the 855 households included in this study, 35% (N=296) reported living on a rural farming property and 40% (N=340) reported living on a rural non-farming property. The remaining households were classified by residents as being a cottage property (3%, N=24), or in either a village or hamlet (13%, N=112), or in a small town (7%, N=57). There was no significant difference between contaminated and uncontaminated wells and the property designation.

The majority of households (75%, N=643) reported that no manure was stored on the property in the 12 months prior to the well water sample being submitted. Two percent (N=15) reported storing it within 15m of the well, 3% (N=25) reported storing it within 30m of the well, 4% (N=38) reported storing it within 100m of the well, 9% (N=78) reported storing it over 100m from the well and 3% (N=27) either did not know if manure was stored or they refused to answer the question. A higher proportion of wells contaminated with *E. coli* were used by households that reported either not knowing if manure was stored or that refused to answer the question about manure storage on their property (5%) compared with wells that were not contaminated (1%).

The majority of households (98%, N=829) reported not spreading sludge from human waste within ninety meters of the well in the year prior to the water sample being collected. Only sixteen households (2%) reported spreading sludge. There was no significant difference between contaminated and uncontaminated wells and spreading sludge from human waste within ninety meters of the well in the 12 months prior to the water sample being collected.

Of the 855 households, 68% (N=584) reported that no livestock were housed on the property in the 12 months prior to water sample collection and 32% (N=271) did. A higher proportion of *E. coli*-contaminated wells were associated with households that reported housing livestock on the property in the 12 months prior to water sample collection (36.1%) compared with wells that were not contaminated (24.3%). Households that reported housing animals on their property were asked to list the types and numbers of animals housed. One-hundred and fifty four (18%) reported housing cattle, which included beef and dairy cattle. Thirty-three (4%) reported housing ruminants, which included sheep and goats. Sixty (7%) reported housing poultry, which included turkeys and chickens. Seventeen (2%) reported housing pigs. One-hundred and nineteen (14%) reported housing horses. Thirty-eight (4%) reported housing other types of animals not noted above, including rabbits, ducks, mink, deer, alpaca and llamas. In comparison to wells that were not contaminated with *E. coli*, a higher proportion of *E. coli*-contaminated wells were associated with households that reported housing cattle (21.0% vs 13.1%), poultry (8.4% vs 4.7%), pigs (3.2% vs 0) or horses (16.5% vs 9.7%,) .

A measure of livestock density, defined as the number of animal units per square kilometer, was also analyzed for association with *E. coli* contamination of the well water. Of the 855 households included in this study, 71% (N=605) had a low animal density, 14% (N=117) had a medium animal density and 15% (N=133) had a high animal density. A higher proportion of *E. coli*-contaminated wells were associated with households that had medium (16.3%) or high (17.0%) animal density compared with wells that were not contaminated (medium animal density=9.4%, high animal density=13.1%).

Table 3. 6: Comparison of identified land use risk factors for the study households using private drinking water wells in Southern Ontario & Southern Alberta (2005-2006)

Variable	All Households N=855 (% Households)	Cases N=534 (% Cases)	Controls N=321 (% Controls)	p-value
Property Designation				
Farm	296 (34.6)	204 (38.2)	92 (28.7)	0.056
Non-farm	340 (39.8)	205 (38.4)	135 (42.1)	
Village or hamlet (< 1000 people)	112 (13.1)	61 (11.4)	51 (15.9)	
Small town (1000 - 10 000 people)	57 (6.7	35 (6.6)	22 (6.9)	
Cottage	24 (2.8)	16 (3.0)	8 (2.5)	
Not Stated	26 (3.0	13 (2.4)	13 (4.1)	
Manure storage in the past year (distance to well, m)				
Within 15m	15 (1.6)	10 (1.9)	5 (1.6)	0.012
Within 30m	25 (2.9)	20 (3.8)	5 (1.6)	
Within 100m	38 (4.4)	21 (3.9)	17 (5.3)	
More than 100m	78 (9.1)	52 (9.7)	26 (8.1)	
Not spread	643 (75.2)	385 (72.1)	258 (80.4)	
Don't know or refused	27 (3.2)	24 (4.5)	3 (0.9)	
Not stated	29 (3.4)	22 (4.1)	7 (2.2)	
Sludge spread within 90m of well in past year				
Yes	16 (1.9)	10 (1.9)	6 (1.9)	0.983
No	829 (97.0)	516 (96.6)	313 (97.5)	
Not stated	10 (1.2)	8 (1.5)	2 (0.6)	
Livestock housed on property in past 12 months				
Yes	271 (31.7)	193 (36.1)	78 (24.3)	<0.001
No	584 (68.3)	341 (63.9)	243 (75.7)	

Table 3.6 continued...

Variable	All Households N=855 (% Households)	Cases N=534 (% Cases)	Controls N=321 (% Controls)	p-value
Type of livestock housed on property in past 12 months				
Cattle	154 (18.0)	112 (21.0)	42 (13.1)	<0.001
No	701 (82.0)	422 (79.0)	279 (86.9)	
Ruminant	33 (3.9)	23 (4.3)	10 (3.1)	0.381
No	822 (96.1)	511 (95.7)	311 (96.9)	
Poultry	60 (7.0)	45 (8.4)	15 (4.7)	0.037
No	795 (93.0)	489 (91.6)	306 (95.3)	
Pig	17 (2.0)	17 (3.2)	0	<0.001
No	838 (98.0)	517 (96.8)	321 (100)	
Horses	119 (13.9)	88 (16.5)	31 (9.7)	<0.001
No	736 (86.1)	446 (83.5)	290 (90.3)	
Other	38 (4.4)	22 (4.1)	16 (5.0)	0.552
No	817 (95.6)	512 (95.9)	305 (95.0)	
Livestock Density (Animal units/km)				
Low (<3)	605 (70.8)	356 (66.7)	249 (77.6)	<0.001
Medium (3-80)	117 (13.7)	87 (16.3)	30 (9.4)	
High (>80)	133 (15.6)	91 (17.0)	42 (13.1)	

The odds ratios for dichotomous and categorical variables identified as land use risk factors investigated are presented in Table 3.7. While overall there was no significant association between property designation and *E. coli* contamination of the water well, there was a significant difference between the reference category, households located on properties reported as a rural non-farms, and households located on properties reported to be farms (OR 1.46 [95% CI, 1.05-2.03]). The odds of a water sample submitted from a well located on a farming property being contaminated with *E. coli* was 1.5 times higher when compared to wells that were on a rural non-farming property.

The storage of manure on the property was associated *E. coli* contamination of the well water (p-value = 0.005). Of households that reported storing manure on their property (N=156), the odds of contamination was highest in households that stored the manure within thirty meters of the well head, in comparison to the reference category, households that had not stored manure on their property in the twelve months prior to the water sample being collected (N=643) (OR 2.68 [95% CI, 0.993-7.23]). The odds of a water sample submitted from a well that had manure storage within 30m being contaminated with *E. coli* was almost 3 times higher when compared to wells from properties where no manure storage was reported. However, only 20 cases and 5 controls reported manure storage within 30m of the well, making it difficult to generalize to the population given the small cell size. This is in line with research previously reported, which stated that the control of animal wastes generated by agricultural activity was identified as a potential non-point source contaminant for drinking water (George *et al.*, 2004, Ritter *et al.*, 2002). For the rest of the households that reported storing manure on their property, the odds of contamination did not vary between households that stored manure within fifteen meters of the well, within one hundred meters of the well, or more than one hundred meters from the well when compared to the reference group. Of interest, when looking at all 855 households included in this research for their response to the question about manure storage, the largest association between *E. coli* contamination of the well water and manure storage was in households that could not remember if manure had been stored or households that refused to answer the question. So while the odds of a water sample submitted from a well that had manure storage within 30 meters being contaminated with *E. coli* was almost 3 times higher when compared to wells from properties where no manure storage was reported; the odds of a water sample submitted from a well used by a households that either reported not knowing whether manure was spread, or refused to answer the question was 5 times higher when compared to households that did not spread manure in the twelve months prior to the water sample being collected (OR 5.36 [95% CI, 1.60-17.99]). Once again, as only 24 cases and 3 controls refused to answer the question or could not remember if manure was stored, it is difficult to generalize these findings to the population given then small cell sizes. In addition, well water contamination from manure storage is more likely associated with the design and construction of the manure storage facilities versus just the presence of manure storage on the property. Households included in this research were not asked questions about the design, construction or maintenance of their manure storage facilities.

Housing livestock on the property anytime in the twelve months prior to collecting the water sample was significantly associated with *E. coli* contamination of the well water (OR, 1.76 [95% CI, 1.29-2.40]). The odds of a water sample submitted from a household with cattle (OR,1.76 [95% CI,1.20-2.59]), horses (OR, 1.84 [95% CI, 1.19-

2.85]), or poultry (OR, 1.88 [95% CI, 1.03-3.43]) being contaminated with *E. coli* was almost two times higher when compared to households that reported housing no animals. The odds were highest when the animals housed were pigs (OR, 14.86 [95% CI, 2.53-inf]). Given these animals are also know to be reservoirs for *E. coli* O157:H7 and other bacterial pathogens such as *Salmonella* and *Campylobacter* species, this reinforces the importance of this study in identifying environmental risk factors associated with the contamination of private drinking water wells (La Ragione *et al.*, 2009, Hancock *et al.*, 2001). Besides the type of animals present, the density of the livestock housed on the property was also significantly associated with water contamination with *E. coli*. For livestock density there was a significant difference between the reference group, low animal density (livestock density <3), and medium animal density (livestock density 3 to 80) (OR 2.03 [95% CI, 1.30-3.17]) as well as high animal density (livestock density >80) (OR 1.52 [95% CI, 1.02-2.26]). The odds of a water sample submitted from a household with medium or high animal density being contaminated with *E. coli* was two times higher when compared to households with a low animal density. This is similar to previous research which has shown different measures of livestock density to be important predictors of risk of *E. coli* contamination of private drinking water (Jokinen *et al.*, 2012; Gannon *et al.*, 2004; Valcour *et al.*, 2002; Michel *et al.*, 1999).

Lastly, there was no significant association between the application of sludge from human waste within ninety meters of the well and *E. coli* contamination of the water well. While there is a risk of environmental contamination from both horizontal and vertical movement of sewage sludge-borne bacteria (Horswell *et al.*, 2010), correct application of biosolids has been shown to have minimal impact on ground water quality.

A study on the impact of long-term land application of biosolids on groundwater and soil quality of an application site operated for 8 to 15 years found bacteriological levels in groundwater samples to be close to background levels for fecal coliforms and below permissible limits (Surampalli *et al.*, 2008). The risk of well water contamination from the application of sewage sludge is more likely associated with how, where and when the biosoilds are applied to the land versus whether they have been applied or not. Households included in this research were not asked for details about the application of biosolids on their property.

Table 3. 7: Bivariate associations between E. coli contamination of ground water used as a private drinking water source and covariates identified as land use risk factors in the survey of households using private drinking water wells in Southern Ontario & Southern Alberta (2005-2006)

Variable	Odds Ratio	95% CI	p-value
Property Designation (ref: Non-farm)			
Farm	1.46	1.05-2.03	0.024
Village or hamlet (< 1000 people)	0.79	0.51-1.21	0.277
Small town (1000 - 10 000 people)	1.05	0.59-1.86	0.874
Cottage	1.31	0.55-3.16	0.538
	Overall	-	0.056
Manure storage in the past 12 months (ref: Not spread in past 12 months)			
Within 15m (50') of well	1.34	0.45-3.97	0.597
Within 30m (100') of well	2.68	0.99-7.23	0.052
Within 100m (330') of well	0.83	0.43-1.60	0.574
More than 100m (330') from well	1.34	0.82-2.20	0.248
Don't know or refused	5.36	1.59-17.99	<0.01
	Overall	-	<0.01
Livestock housed on property in past 12 months (ref: No)			
Yes	1.76	1.29-2.40	<0.001

Table 3.7 continued...

Variable	Odds Ratio	95% CI	p-value
Type of livestock housed on property in past 12 months (ref: No)			
Cattle	1.76	1.20-2.59	<0.001
Ruminant	1.40	0.66-2.98	0.383
Poultry	1.88	1.02-3.43	0.040
Pigs*	14.87	2.54-inf.	<0.01
Horses	1.84	1.19-2.85	<0.01
Other	0.82	0.42-1.58	0.553
Livestock Density (ref: Low)			
Medium	2.03	1.30-3.17	<0.001
High	1.52	1.02-2.26	0.042
Sludge from human waste spread within 90m of well in past 12 months (ref: No)			
Yes	1.01	0.36-2.80	0.983

* Exact logistic regression used to calculate OR.

3.2.4 Temporal Risk Factors – Results and Discussion

Table 3.8 outlines the different temporal risk factors that were analyzed for association with *E. coli* contamination of the well water. The majority (61%, N=527) of households included in this research submitted their well water sample for testing during the summer months, defined as June to the end of August. Of the remaining households, 17% (N=146) submitted their water sample during the fall, defined as September to the end of November, 14% (N=123) submitted their water sample during the spring, defined as March to the end of May, and 7% (N=59) submitted their water sample during the winter, defined as December to the end of February. There was no significant difference between contaminated and uncontaminated wells and the season in which the water sample was submitted for testing.

Of the 855 households included in this study, 65.5% (n=560) had no rainfall on the day the sample was collected, 63.6% (n=544) had no rainfall 1 day prior to the collection of the water sample, 59.9% (n=512) had no rainfall 2 days prior to the collection of the water sample and 61.3% (n=532) had no rainfall 3 days prior to the collection of the water sample. There was no significant difference between contaminated and uncontaminated wells and rainfall on the day the water sample was collected (p-value = 0.794), nor rainfall one day (p-value = 0.685), two days (p-value=0.860) or three days (p-value=0.406) prior to the water sample collection. The total rainfall over the four days prior to sample collection ranged between 0mm up to 126mm with a median amount of 4mm (mean=10mm). There was no difference between the contaminated and uncontaminated wells and the total rainfall over the four days prior to water sample collection (p-value = 0.532).

Table 3. 8: Comparison of identified temporal risk factors for the study households using private drinking water wells in Southern Ontario & Southern Alberta (2005-2006)

Variable	All Households N=855 (% Households)	Cases N=534 (% Cases)	Controls N=321 (% Controls)	p-value
Season Water Sample Taken				
Spring (Mar to May)	123 (14.4)	68 (12.7)	55 (17.1)	0.131
Summer (Jun to Aug)	527 (61.6)	337 (63.1)	190 (59.2)	
Fall (Sept to Nov)	146 (17.1)	87 (16.3)	59 (18.4)	
Winter (Dec to Feb)	59 (6.9)	42 (7.9)	17 (5.3)	
Rainfall on day of sample (mm)				
0	560 (65.5)	348 (65.2)	212 (66.0)	0.794
>0	372 (43.5)	186 (34.8)	186 (57.9)	
Rainfall 1 day prior to sample (mm)				
0	544 (63.6)	337 (63.1)	207 (64.5)	0.685
>0	311 (36.4)	197 (36.9)	114 (35.5)	
Rainfall 2 day prior to sample (mm)				
0	512 (59.9)	321 (60.1)	191 (59.5)	0.860
>0	343 (40.1)	213 (39.9)	130 (40.5)	
Rainfall 3 day prior to sample (mm)				
0	524 (61.3)	333 (62.4)	191 (59.5)	0.406
>0	331 (38.7)	201 (37.6)	130 (40.5)	
Total Rainfall (mm)				
min to 4	423 (49.5	254 (47.6)	169 (52.7)	0.532
4.1 to 8	116 (13.6)	75 (14.0)	41 (12.8)	
8.1 to 12	98 (11.5)	65 (12.2)	33 (10.3)	
12.1 to max	218 (25.5)	140 (26.2)	78 (24.3)	

The odds ratios for dichotomous and categorical variables identified as temporal factors investigated are presented in Table 3.9. For the season in which the water sample was collected and *E. coli* contamination of the water well, while there was a significant difference between samples collected during the winter months and the reference category, samples collected during the spring. This division is likely a product of the selection of households for the case-control study completed as part of the original research project: controls were matched to cases within ±1 month of water collection.

Previous research has identified different climatic factors that act as potential contributors to well water contamination events. Richardson *et al.* (2009) found that the likelihood of well failure increased with increasing intensity of rainfall on the day prior to sampling, with *E. coli* being more likely to be isolated in the water samples if they were obtained one day after a heavy rainfall (defined as greater than 1000mm). A study conducted by Jokinen *et al.* (2012) in the Old Man River Watershed in Southern Alberta found that the sum total of rainfall over the course of the three days prior to the sample being taken was the best classifier of *E. coli* O157:H7 presence surface water sources. For the research carried out in this thesis, there was no significant association between any of the rainfall data collected and *E. coli* contamination of the water well. However, the rainfall data was collected from the nearest station and may not accurately represent the amount that fell on the property.

Table 3. 9: Bivariate associations between E. coli contamination of ground water used as a private drinking water source and covariates identified as temporal risk factors in the survey of households using private drinking water wells in Southern Ontario & Southern Alberta (2005-2006)

Variable	Odds Ratio	95% CI	p-value
Season Water Sample Taken (ref: Spring (Mar to May))			
Summer (Jun to Aug)	1.43	0.96-2.13	0.075
Fall (Sept to Nov)	1.19	0.73-1.94	0.477
Winter (Dec to Feb)	2.00	1.03-3.89	0.042
	Overall	-	0.130
Rainfall on day of sample (mm) (ref: 0)			
>0	1.04	0.77-1.39	0.794
Rainfall 1 day prior to sample (mm) (ref: 0)			
>0	1.06	0.80-1.42	0.685
Rainfall 2 day prior to sample (mm) (ref: 0)			
>0	0.97	0.74-1.29	0.860
Rainfall 3 day prior to sample (mm) (ref: 0)			
>0	0.87	0.67-1.18	0.406
Total Rainfall (mm) (ref: min to 4)			
4.1 to 8	1.22	0.79-1.87	0.369
8.1 to 12	1.31	0.83-2.08	0.251
12.1 to max	1.19	0.85-1.68	0.304
	Overall	-	0.531

3.3 Summary

Based on bivariate analysis alone, several risk factors have been associated with *E. coli* contamination of the water well. The physical characteristics of the well itself were as important as point sources of pollution with regard to the contamination of water wells with *E. coli*. The depth and type of well were shown to play an important role in affecting the vulnerability of a well water source to contamination. The manure storage, housing of livestock along with the type of livestock housed on the property were all identified as significant point sources of contamination for well water. Similarly, the age of the septic system, which may be a proxy for whether the septic system itself was malfunctioning was also identified as an important risk factor for the contamination of well water. These findings are expected given the likelihood of groundwater contamination depends on the intrinsic susceptibility of the well, which is related to the hydrological characteristics of the groundwater source, along with aquifer vulnerability, which is related to the effect of land-use practices and contaminant characteristics and loading (Frind *et al.*, 2006). The development of a robust predictive model on when groundwater contamination events are most likely to occur requires the incorporation of several site-specific parameters. These interactions are further investigated in Chapter 4.

CHAPTER 4

An Evaluation of Environmental Factors Affecting Microbial Contamination of Drinking Water Wells Using Multivariable Analysis

4.1 Introduction

To date, few studies using epidemiological data to investigate the effect that multiple environmental risk factors play in facilitating contamination of groundwater have been conducted. The results of the few studies completed in the United States have focused on the association between different risk factors and well water contamination with total coliforms and *E. coli*, and have shown mixed results (Allevi *et al.*, 2013; Won *et al.*, 2013; Swistock *et al.*, 2013).

The objective of the research presented in this chapter was to detail the results of the multivariable analysis, identifying significant environmental factors that were associated with the contamination of private drinking water wells with *E. coli*.

There were several sets of potential mediators investigated for this study, as outlined *a priori* and depicted in Figure 1 (Section 2.4 – Variable Identification). Mediation occurs when a third variable carries the influence of a given independent variable (Frazier *et al.*, 2004). Mediators establish how or why one variable predicts or causes an outcome variable by acting as the mechanism through which an independent variable influences the dependent variable (Frazier *et al.*, 2004; Baron *et al.*, 1986). In addition to the designation of variables as confounders, the designation of mediators was used as a tool to describe how the final variables in the multivariable model interacted with each other.

Based on the results of the bivariate analysis or as a result of their environmental significance, variables were selected for inclusion into the full model. Prior to building the final model, assessment for collinearity was carried out using the Cramer's V statistic. Where variables were identified as collinear, the most biologically relevant variable was included and the other collinear variable dropped. The final model consisted of those variables significant at a p-value of 0.05, identified mediators, confounders and any significant interactions.

4.2 Results and Discussion

4.2.1 Assessment for Mediation

The mediation effect of the identified potential mediators was carried out using the strategy outlined by Baron *et al.* (1986), as reviewed in Chapter 2 of this thesis. Verification of mediation is based on the finding that if the mediator were to be removed, the relationship between the independent and dependent variables would be significantly changed.

4.2.1.1 Mediation of On-site Waste Water Control Systems

The type of on-site wastewater control system used on a property was identified as a potential independent variable for the dependent variable (i.e., contamination of the well water on the property with *E. coli*). Potential mediators for this relationship included the distance between the wellhead and the septic tank, the distance between the well head and the weeping tile as well as the age of the septic system.

As previously discussed in Chapter 3 (Section 3.2 and Tables 3 and 4), the type of on-site wastewater treatment system used was not a significant predictor of *E. coli* contamination of the well water on the property. To prove mediation, Baron requires that the independent variable be shown to affect the dependent variable. Given this initial requirement was not met; no further assessment for mediation of the impact of on-site wastewater treatment systems used was warranted.

4.2.1.2 Mediation of the Type of Water Well

The type of well used to supply the drinking water for the property was identified as a potential independent variable for contamination of the well water on the property with *E. coli*. Potential mediators for this relationship included the depth of the well, the age of the well, the total rainfall that fell over the four days prior to the water sample being collected and the season in which the water sample was taken.

The age of the well, when the well was greater than 25 years, remained significant after controlling for well type (p-value for well age 25 to 35 years was 0.003; p-value for well age greater than 35 years was 0.006), thus the designation as a mediator was supported for the age of the well.

Neither the depth of the well, total rainfall over the four days prior to the collection of the water sample, nor the season in which the water sample was collected remained significant after controlling for well type. A summary of the p-values can be found in Table 4.1 below. The designation as mediator was not supported for these variables.

Table 4. 1 Multivariable associations between contamination of well water with E. coli and identification of potential mediators when controlling for well type using logistic regression

Mediating Variable		**p-value**
Well Depth (m)	Less than 10	Ref
	10 to 30	0.566
	30.1 to 40	0.321
	Greater than 40	0.130
Total Rainfall day of and three days prior to sample collection (mm)	Less than 4	Ref
	4 to 8	0.369
	8.1 to 12	0.251
	Greater than 12	0.304
Season	Spring (March to May)	Ref
	Summer (June to August)	0.120
	Fall (September to November)	0.730
	Winter (December to February)	0.103

4.2.1.3 Mediation of Livestock Housed on the Property

The housing of livestock on the property in the twelve months prior to the water sample being collected was identified as a potential independent variable for the dependant variable (i.e., contamination of the well water on the property with *E. coli*). Potential mediators for this relationship included the type of livestock housed and the livestock density for the property.

When assessing the impact of the different types of livestock housed on a property (potential mediator variables) on livestock housed on the property in the 12 months prior to the water sample collection (independent variable), it was determined that the potential mediator variables (the type of livestock housed) separated the independent variable (livestock housed in the 12 months prior to water sample collection) completely. For all instances where the type of livestock housed was answered yes (e.g., cattle housed on property = yes), livestock housed on the property in the 12 months prior to water sample collection was also yes. This type of problem with the data is known as separation in logistic regression and occurs when the responses and non-responses can be perfectly separated by a single risk factor (Heinze *et al.*, 2002). As the type of livestock housed on the property and the housing of livestock on the property were measuring the same concept, the designation as mediators for the types of livestock housed on the property was not supported, and the type of livestock housed was dropped from further analysis.

Lastly, livestock density (p-value for medium animal density 0.778; p-value for high animal density 0.782) did not remain significant after controlling for the housing of livestock on the property in the twelve months prior to the water sample being collected, thus mediation was not supported for this variable.

4.2.1.4 Mediation Assessment Summary

After completing the assessment for mediation, the age of the well was found to be a mediator for the relationship between the type of well and well water contamination with *E. coli*. No other mediating relationships were identified, but the variables assessed as potential mediators were still considered, based on bivariate analysis with the dependent variable (well water contamination with *E. coli*), for inclusion in the full model as potential confounders for the identified independent variables. This is discussed in further detail in section 4.4 of this chapter (Association of Environmental Risk Factors and Contaminated Well Water).

4.3 Assessment for Collinearity

The depth of the well and the type of the well were identified as collinear (Cramer's V=0.5149). As mentioned in Chapter 1, section 1.2, well depth is very closely tied to the type of well being established. Dug wells are shallow holes that are dug approximately 3 to 9 meters deep, driven wells are created when pipe is driven through gravel or sandy soil approximately 15 meters deep, and drilled wells are wells that are drilled until they reach bedrock which is typically at depths between 30 to 120 meters (Committee on Environmental Heath and Committee on Infectious Disease, 2009). Given that the depth of the well is dictated by the type of well being installed at the site well depth was not included in the development of the final model and the type of well was retained.

The storage of manure on the property, and the housing of livestock on the property were found to be collinear (Cramer's V=0.5734). As well, animal density and the housing of livestock on the property were found to be collinear (Cramer's V=0.9437). The housing of livestock on the property was the most direct measure of the impact of livestock on a private well when compared to the storage of manure on the property and animal density. Manure storage on the property and animal density were not included in the development of the final model and the housing of livestock on the property was retained.

4.4 Association of Environmental Risk Factors and Contaminated Well Water

After removing identified collinear variables, remaining variables that were associated with contamination of the well water with *E. coli*, at a p-value of 0.25 or less on bivariate analyses, were used to build the multivariable model (see Appendix 3) using purposeful selection. The initial, or full, model included the following variables: independent variables - the type of the water well and livestock housed on the property in the 12 months prior to water sample collection; potential confounders - total income, education, province, mode of interview, soil type and property designation; and the mediator - age of the well. While mediation analysis did not support including the age of the septic system, total rainfall in the three days prior and the day of the water sample collection or the season in which the water sample was collected as mediators, these variables were significant at the p-value of 0.25 or less on bivariate analysis with the dependent variable, as such, included in the full model for assessment.

Variables with a p-value greater than 0.25 on bivariate analysis that were not considered environmentally significant, were not included in the multivariable model. This included, the rainfall on the day the well water sample was collected (p-value=0.794), rainfall 1 day prior to the well water sample being collected (p-value=0.685), rainfall 2 days prior to the well water sample being collected (p-value=0.860) and rainfall 3 days prior to the well water sample being collected (p-value=0.406). In addition, they type of on-site waste water management system used (p-value=0.874), the distance in m between the well head and the weeping tile (p-value=0.789), the distance in m between the well head and the septic tank (p-value=0.511) and the spreading of sludge from human waste within 90m of the well head (p-value=0.983). Lastly, as discussed above (section 4.2, Mediation of Livestock Housed on the Property), the housing of livestock on the property and the type of livestock housed were both measuring the same concept. Including both measures would have been redundant. The housing of livestock on the property in the 12 months prior to the water sample being collected was identified as one of the main independent variables for this research; as such the types of livestock housed were dropped.

Variables from the multivariable model with a p-value ≤ 0.05 were retained for the main effects model. Prior to dropping variables from the main effects model, the confounding effects of the variables to be removed were assessed by determining whether the change in the regression coefficients of the primary predictors, which stayed in the reduced model, changed by more than 15% when the variables were removed (Hosmer *et al.*, 2000). If any of the regression coefficients changed by more than 15%, the removed variable was included back in the model. This process was continued until the important variables were included in the model and those excluded were statistically and/or environmentally unimportant (Hosmer *et al.*, 2000). After eliminating variables based on these conditions, the main effects model included the independent variables: type of the water well, livestock housed on the property in the 12 months prior to water sample collection; the confounders: total income, education and soil type; and the mediator: age of the well.

Once the main effects model was complete, biologically plausible interactions among variables in the main effects model were assessed. Interaction variables were created as the arithmetic product of the pairs of main effect variables (Hosmer *et al.*, 2000). The potential interaction variables were added into the model one at a time. Any interaction variables that did not remain significant when included in the model were dropped. The potential interactions assessed included: manure storage on the property in the 12 months prior to water sample collection and livestock housed on the property in the 12 months prior to water sample collection; livestock housed on the property in the 12 months prior to water sample collection and the type of well; and, depth of well and age of well, were entered into the final model. None of the product terms were significant, and thus were removed from the model. The final model included the independent variables: type of the water well, livestock housed on the property in the 12 months prior to water sample collection; the confounders: total income, education and soil type; and the mediator: age of the well (see Table 4.2).

The final model used 821 observations. The Hosmer-Lemshow goodness of fit test was not significant (p-value = 0.999), indicating that the model fit well (Hosmer *et al.*, 2000). In addition, the linktest, used to detect specification error, was not significant, indicating that there were no additional predictors that were statistically significant, supporting the conclusion that the model was properly specified.

After adjusting for the effect of the other variables, there was a significant difference between the type of well used and the odds of *E. coli* contamination. When compared to the reference group, households that used a drilled well, the odds of a water samples submitted from dug or bored well being contaminated with *E. coli* was over two times higher (OR 2.35 [95% CI,1.60-3.46]). The odds of a water sample submitted from a driven or sandpoint well being contaminated with *E. coli* was only 1/3 when compared to drilled wells (OR 0.41 [95% CI, 0.16-1.09]).

For the age of the well, there was no significant difference between the reference group, well age less than ten years, and households with a well between ten and twenty five years of age. After adjusting for the effect of the other variables, the odds of *E. coli* contamination was significant for households with a well that was between 26 years to 35 years old (OR 1.88 [95% CI, 1.23-2.87]) and wells that were greater than 36 years of age (OR1.79 [95% CI, 1.12-2.86]). The odds of a water sample submitted from a well that was 26 years or older being contaminated with *E. coli* was almost two times higher when compared to wells that were less than ten years old.

Lastly, the housing of livestock on the property in the 12 months prior to the water sample collection was also significantly associated with well water contamination with *E. coli*, after controlling for the effects of the other variables. The odds of well water contamination with *E. coli* for households with livestock housed on the property was over 2 times higher when compared to households that housed no livestock in the 12 months prior to water sample collection (OR 2.22 [95% CI 1.57-3.14]).

The confounders included soil type, education and total income. After adjusting for the effect of the other variables, neither the association between the predominant soil type reported for the property nor the highest level of education achieved in the household were significantly associated with *E. coli* contamination of the well water. These variables were retained in the final model as a result of their confounding effect on the main independent variables. For total income, while there was no association between the reference category, households with a total income of less than $40,000, and households with a total income above $60,000, the odds of well water contamination with *E. coli* for households with a total income of between $40,000 and $60,000 was almost 2 times higher when compared to the reference category (OR 1.74 [95% CI 1.02-2.98]).

The results of this multivariable analysis are similar to findings reported by Swistock *et al.* (2013). Their study focused on private drinking water wells used in Pennsylvania and the association between identified risk factors and the presence of both coliform bacteria and *E. coli*. Using multivariable analysis the authors found statistically significant associations between the bedrock geology, soil moisture conditions for the 2 week period prior to sample collection, and well score, a score created for the presence/absence of 5 important well construction characteristics (1-the presence of a metal or plastic casing on the water well, 2-water well casing that extended above ground or was entirely buried, 3-visible evidence of grout or cement around the water well casing, 4-ground slop that promoted the movement of surface water toward or away from the water well casing and 5-the presence of a well cap) and the presence of both coliform bacteria and *E. coli* (Swistock *et al.*, 2013). Similarly, the research presented in this study supported an association between the type of well used and *E. coli* contamination. The important well construction characteristics identified by Swistock *et al.* (2013) as providing protection are similar to the characteristics of the definition of a drilled well in this study. While Swistock *et al.* (2013) did not find well age to be important in explaining water quality, the important well construction characteristics they did identify, such as the use of well casing, the evidence of grout or cement around the casing and the extension of the casing above the ground, would be more common in recently installed wells. The mediating effect identified for well age in this research indicated that part of the association between well type and contamination with *E. coli* is mediated by well age. A component of the relationship between well age and *E. coli* contamination identified in this study was likely due to the fact that older wells are less likely to have been constructed using the methods Swistock *et al.* (2013) identified as protective. In addition, soil type was identified as a confounder in this study, but unlike Swistock *et al.* (2013), after controlling for the other variables, there was no significant association between *E. coli* contamination of the well and the soil type reported.

A study carried out by Allevi *et al.* (2013) on private wells used for drinking water in Virginia developed a final regression model that included 3 significant predictors of total coliform contamination, well depth, whether there was any type of water treatment device and whether the well was located within 1km of a farm animal operation. Although Allevi *et al.* (2013) examined the relationship between contamination with total coliforms, and this study focused on *E. coli*, both microbial populations are used as indicators of microbial quality and safety when assessing water samples from private wells used as a source of drinking water. Allevi *et al.* (2013) found well depth to be significantly associated with total coliform contamination. Similarly, the research for this thesis showed an association between well depth and well water contamination with *E. coli* using bivariate analysis, because well depth and well type were identified as collinear, well depth was not included in the final model for assessment. That being said, as reported in Chapter 3 (Section 3.3 – Physical Risk Factors), well depth is closely related to well type, with the depth of the well being dictated by the type of well installed. In addition, Allevi *et al.* (2013) found an association between total coliform contamination of the well water and the location of a farm animal operation within 1km of the well. Similarly, this research showed an association between *E. coli* contamination of the well water and the presence of farm animals on the property following the multivariable analysis. After controlling for the effects of the other variables, the odds of well water contamination with *E. coli* for households that housed livestock on the property was over 2 times higher when compared to households with no livestock.

On the other hand, a study carried out on private wells used for drinking water in Northeastern Ohio found that the probability of detecting total coliforms was not associated with well depth or age (Won *et al.*, 2013). Won *et al.* (2013) stated that the physical factors, well depth, age and location of the well, which are typically used to identify characteristics of wells at high risk for contamination were not predictive of the likelihood of contamination with microbiological indicators or *E. coli* O157:H7 or *Campylobacter spp*. The authors felt instead that contamination may be associated with other factors which well owners have greater control such as well maintenance. The sample size for their study was 180 randomly selected wells in 2 northeastern Ohio counties. Of the wells selected, *E. coli* was present in only 16 of the 180 samples. This relatively small number of *E. coli* positive water wells may have impacted the ability to identify associations between contamination with *E. coli* and the physical risk factors analyzed.

In addition to the direct associations identified between *E. coli* contamination of private drinking water wells and the type and age of the well installed, a mediating relationship was also established between the type of well installed, the age of the well and *E. coli* contamination. This is the first time mediation has ever been explicitly used to examine the relationships between environmental risk factors and *E. coli* contamination of private drinking water wells. As discussed in section 4.2 above, the age of the well was identified as a mediator in the relationship between the type of well installed and *E. coli* contamination of the water well. Mediation implies a causal sequence among the 3 variables, with the independent variable causing the mediator and the mediator causing the dependent variable. This relationship is different than an interaction effect where an interaction means the effect of the independent variable on the dependent variable depends on the level of the third variable. No causal sequence is implied by interaction.

The identification of the mediation effect of well age means that a proportion of the effect that well type has on *E. coli* contamination is mediated by the age of the well. Once again, to demonstrate mediation, the independent variable must be shown to affect the outcome, the independent variable must be shown to affect the mediator and the mediator must be shown to affect the outcome when controlling for the independent variable. The statistical support for the designation of well age as a mediator for well type is discussed in section 4.2 above. In turn the explanation of the mediation effect is shown through the following relationships. In Chapter 3, section 3.3, which identified physical risk factors associated with *E. coli* contamination of drinking water wells, the relationship between the type of well installed and *E. coli* contamination was already established. Secondly, in this study, the 34% of dug or bored wells were between 25 and 35 years, while only 13% of dug or bored wells were less than 10 years old. This is likely due in part that historically, dug or bored wells would have been more commonly installed, but as information about well safety and the association between contamination likelihood and the type of well used increased (Audette *et al.*, 2001; Goss *et al.*, 1998), the installation of these well types likely decreased. Given the use of dug or bored wells 25 to 35 years ago would have been more likely, there is support for the second component of the mediating relationship between well type and *E. coli* contamination by well age. That is the independent variable is shown to affect the mediator. Lastly, as discussed above, the relationship between well age and *E. coli* contamination, when controlling for the other significant risk factors, has also been established, supporting the final component of the mediating relationship between well type and *E. coli* contamination by well age. That is the mediator is shown to affect the outcome variable when controlling for the independent variable.

Table 4. 2: Bivariate and multivariable model of association between E. coli contamination of ground water used as a private drinking water source and significant risk factors and confounders in the survey of households using private drinking water wells in Southern Ontario & Southern Alberta (2005-2006)

Variable	Bivariate			Multivariable		
	Odds Ratio	**95%CI**	**p-value**	**Odds Ratio**	**95%CI**	**p-value**
Well Type (ref: Drilled)						
Dug or Bored	2.41	1.68-3.45	<0.001	2.35	1.60-3.46	<0.001
Sandpoint	0.33	0.13-0.82	0.017	0.41	0.16-1.09	0.075
Well Age (years) (ref: min to 10)						
10.1 to 25	1.19	0.82-1.74	1.19	1.10	0.73-1.64	0.653
25.1 to 35	2.08	1.39-3.08	2.08	1.88	1.23-2.87	<0.01
35.1 to max	2.19	1.42-3.38	2.19	1.79	1.12-2.86	0.015
Total Household Income (ref: Less than $20 000 to less than $40 000)						
$40 000 to less than $60 000				1.74	1.02-2.98	0.042
$60 000 to less than $80 000				0.89	0.53-1.51	0.675
$80 000 or more				1.22	0.74-2.03	0.433
Don't know or refused or not stated				0.87	0.60-1.26	0.466
Soil Type (ref: Sand)						
Gravel	1.44	0.90-2.29	0.129	1.44	0.86-2.42	0.162
Loam	1.13	0.75-1.69	0.572	0.86	0.55-1.36	0.529
Clay	1.58	1.06-2.35	0.024	1.34	0.87-2.07	0.187
Livestock Housed on the Property in the past 12 Months (ref: No)						
Yes	1.76	1.29-2.40	<0.001	2.22	1.57-3.14	<0.001

Table 4.2 continued...

Highest Education in Household (ref: Graduated High School)						
Less than grade 9 or Some high school	1.76	0.91-3.37	0.091	1.42	0.72-2.81	0.49
College or trade school	1.19	0.78-1.80	0.417	1.26	0.81-1.96	0.28
University	1.14	0.76-1.69	0.520	1.25	0.81-1.93	0.28

4.5 Summary

The results from the logistic regression analysis demonstrate that the presence of environmental risk factors, in particular the type of well used, the age of the well and the housing of livestock on the property are associated with the contamination of private drinking water supplies in Ontario and Alberta with *E. coli*. In addition to the direct effect that the type of well installed has on *E. coli* contamination, the assessment for mediation has also demonstrated the mediated effect that well type has on *E. coli* contamination through well age. This research provides evidence that identified environmental risk factors are correlated with *E. coli* contamination, in addition the pathways through which those environmental risk factors act to cause contamination are more completely described.

Chapter 5

General Discussion

5.1 Significant Findings

The significant findings in this thesis indicate that several environmental risk factors, most notably, well type, well age and the housing of livestock on a property are significantly associated with *E. coli* contamination of private drinking water wells in Southern Ontario and Alberta. This chapter discusses the relevance, application and impact of these findings.

5.2 Environmental Risk Factors Associated with *E. coli* Contamination of Private Well Water

The overall objective of this study was to identify significant environmental factors that were associated with the contamination of private drinking water wells with *E. coli*. After adjusting for the effect of household income and soil type, the odds of a water sample submitted from dug or bored well being contaminated with *E. coli* was over two times higher when compared to drilled wells. There was no difference in *E. coli* contamination between sandpoint wells and drilled wells. The odds of a water sample submitted from a well that was 25.1 years or older being contaminated with *E. coli* was almost two times higher when compared to wells that were less than ten years old. Lastly, the odds of well water contamination with *E. coli* for households with livestock housed on the property was over 2 times higher when compared to households that housed no livestock in the 12 months prior to water sample collection.

These findings were similar to two studies carried out in the United States, both completed within the last couple of years. One study looked at the association between potential risk factors and the presence of both total coliforms and *E. coli* (Swistock *et al.*, 2013) the other, the association between risk factors and the presence of total coliforms (Allevi *et al.*, 2013). Using multivariable logistic regression, Swistock *et al.* (2013) found that the bedrock geology around the well, the soil moisture over the 2 weeks prior to the sample collection and the well score, which was created for the presence or absence of 5 important well construction characteristics were significant predictors of both total colifom and *E. coli* contamination. Allevi *et al.* (2013) found that well depth, the use of any type of treatment device and the location of the well within 1km of a farm animal operation to be significantly associated with total coliform contamination of the well.

In addition to the direct associations identified between *E. coli* contamination of private drinking water wells and the type and age of the well installed, a mediating relationship was also established between the type of well installed, the age of the well and *E. coli* contamination. To our knowledge, there are currently no other published studies that have looked into identifying environmental risk factors as mediators in private well water contamination with *E. coli*. Distinguishing this relationship provides a clearer understanding of how these variables may interact to impact the susceptibility of private water wells to contamination with *E. coli*.

While the variables for the different types of livestock housed on the property were not included in the multivariable model as a result of separation with the variable for housing livestock on the property, the housing of cattle, horses, poultry and pigs were all found to be significantly associated with *E. coli* contamination of the well water in the univariate analysis. The odds of a water sample submitted from a household with cattle, horses, or poultry being contaminated with *E. coli* was almost two times higher when compared to households that reported housing no animals. Of note, the odds were highest when the animals housed were pigs (OR, 14.86 [95% CI, 2.53-inf]). Given these animals are also know to be reservoirs for *E. coli* O157:H7 and other bacterial pathogens such as *Salmonella* and *Campylobacter* species, further investigation into the association between the housing of pigs and *E. coli* contamination of well water may be warranted. (La Ragione *et al.*, 2009, Hancock *et al.*, 2001).

This study provides additional information to previous efforts which correlated private water supply contamination with predictive factors by identifying specific risk factors associated with contamination of private drinking water wells in Southern Ontario and Alberta.

5.3 Strengths and Limitations

This study had a number of strengths and limitations. This study is the first analytic case control study on the identification of environmental risk factors associated with private drinking water well contamination to be completed, that includes households in both the province of Ontario and Alberta.

Another strength included the use of a standardized questionnaire that allowed for questions on a number of different potential confounders and effect modifiers. The answers from these questions were in turn used in logistic regression analysis to control for their effect so that the true nature of the relationship between *E. coli* contamination and private drinking water wells and environmental risk factors could be identified.

The analysis conducted in this study included the assessment for mediation. While hypotheses regarding mediated effects are common in psychological research (MacKinnon *et al.*, 2000), this study represents the first time such an analysis has been carried out to explain the association between environmental risk factors associated with *E. coli* contamination of private drinking water wells. The identification of mediators in this study helps to more clearly explain how the significant independent variables impacted the outcome variable, contamination of well water with *E. coli*. In addition, this is the first time multivariable logistic regression analysis was used to investigate potential correlations between *E. coli* contamination of private drinking water wells and environmental risk factors in two provinces in Canada.

As in all research, this study also had limitations. The case-control study was limited by having access only to those households that submitted a water sample to participating public health laboratories. Not all households with private drinking water wells submit their water for bacteriological testing, as such, the sample may not be truly representative of all residents that use private drinking water wells in Ontario and Alberta. However, questionnaires submitted were from wide geographical areas of both Ontario and Alberta, and included farming and non-farming properties. In addition, when well owners were contacted for participation in the study, they would have been aware of the results of their well water test, this could have introduced bias in participation and recall for cases and controls.

There was a lag between when a household submitted a water sample for bacteriological analysis and when the participant was asked about their exposures. This may have resulted in problems with a participant's recall. In addition, data collected on the questionnaires were collected based on self-reports which are less precise than objective measurements. There were two administrative modes of data collection used, personal and telephone interviews. While no significant differences between households that were interviewed via on-site visits and households that were interviewed via telephone were identified, the data collected by different administrative modes might be different in terms of quality and validity. Variables where a measurement was taken by the on-site interviewer may be more accurate than those self-reporting during a phone interview as the interviewer on-site would have measured the distances directly. That being said, none of the variables that included a measurement taken by an on-site interviewer ended up being included in the final multivariable model. In addition, there were no differences in the proportion of case and control participants interviewed by each method, which might have indicated a bias in responses.

The rainfall data collected for this study was collected from the nearest weather station and may not have accurately represented the amount that fell on the property. Rainfall patterns are not evenly dispersed throughout Southern Ontario or Alberta which means rainfall events recorded at weather stations may be very different than what occurred on the properties the wells were located on. As there was no differences observed between the cases and controls, the effect of this limitation was likely minimal.

The data for this thesis was drawn from a larger case control study in which cases and controls were frequency matched by month of sample collection and laboratory region. This matching technique may have resulted in the controls being more similar to the cases than in the source population. Matched pair analysis was not used to analyze the data, which could have biased the results towards the null. However, the designation of cases and controls for this thesis resulted in a subset of the controls from the original study being used as cases in this analysis, negating any matching that may have been used.

Another limitation of this study was that the use of the method outlined by Baron *et al.* (1986) to establish mediation, while valid, does not actually quantify the mediation effect. As such, the mediation effect, while it may have been found to be present, may not be significant.

5.4 Conclusions

The relationships identified in this thesis do suggest that certain private well characteristics and land use practices should be considered when proactively working to prevent or when looking for factors associated with drinking water well contamination. The type of drinking water well installed, when it was installed, and whether livestock was housed on the property were all found to be significantly associated with contamination with E. coli, even when the association was adjusted for household income and soil type. Regulators responsible for water well construction may wish to consider the association between water well type and the occurrence of *E. coli* when writing and/or revising water well guidelines and regulations. In addition, this information can be used by public health inspectors advising private well water owners with respect to potential risk factors associated *E. coli* contamination. Inspectors may wish to use this information to advise owners of the risk of recontamination, the need for the submission of routine water samples for bacteriological analysis, and consideration of permanent treatment systems for their water supply

In 2006, Environment Canada reported approximately 4 million groundwater users lived in rural areas and accessed private drinking water supplies (Environment Canada, 2006). In both Alberta and Ontario, the local health unit is where these private well owners visit to submit their water well samples for bacteriological testing, and to receive assistance in interpreting the results of those tests. Arming public health officials with the significant findings from this thesis will allow them to highlight factors home owners should consider when working to prevent well water contamination and help identify potential sources of pollutants.

References

Allevi, R., Krometis, L., Hagedorn, C., Benham, B., Lawerence, A., Ling, E., Zielgler, P. (2013). Quantitative analysis of microbial contamination in private drinking water supply systems. Journal of Water and Health, 11, 244-255.

Aramini, J., McLean, M., Wilson, J., Holt, J., Copes, R., Allen, B., Sears, W. (2000) Drinking water quality and health-care utilization for gastrointestinal illness in greater Vancouver. Ottawa, Ontario: Health Canada.

Aschengrau, A, Seage, G. (2008). Essentials of epidemiology in public health. Sudbury, Massachusetts: Jones and Bartlett Publishers.

Audette, R., Chanasyk, D., Fitgerald, D., Kiely, R., Neilson, R. (2001). Farm well water quality in Alberta. Water Quality Research Journal of Canada, 36, 565-588.

Baron, R., Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, 51, 1173-1182.

Barrell, R., Hunter, P., Nichols, G. (2000). Microbiological standards for water and their relationship to health risk. Communicable Disease and Public Health, 3, 8-13.

Berry, E., Miller, D. (2005). Cattle feedlot soil moisture and manure content: Impact on *Escherichia coli* O157. Journal of Environmental Quality, 34, 656-663.

Bigras-Poulin, M., Ravel, A., Belanger, D., Michel P. (2004). Development of agroenvironmental indicators to evaluate the hygienic pressure of livestock production on human health. International Journal of Hygiene and Environmental Health, 207, 279-295.

Borchardt, M., Chyou, P., DeVries, E., Belongia, E.. (2003). Septic system density and infectious diarrhea in a defined population of children. Environmental Health Perspectives, 111, 742-748.

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. Journal of Public Health, 27, 281–291.

Brennan, F., O'Flaherty, V., Kramers, G., Grant, J., Richards, K. (2010). Long-term persistence and leaching of *Escherichia coli* in temperate maritime soils. Applied and Environmental Microbiology, 76, 1449–1455.

Catania, J., Oakley, L., Rosen, R., Pollack, L. (2012). Effects of interview mode on assessments of erectile and ejaculatory dysfunction among men with benign prostatic hyperplasia (BPH). Journal of Sex Research, 50, 524-536.

Chalmers, R., Aird, H, Bolton, F. (2000). Waterborne *Escherichia coli* O157. Journal of Applied Microbiology, 88, 124S-132S.

Charrois, Jeffrey W.A. (2010). Private drinking water supplies: challenges for public health. Canadian Medical Association Journal, 182, 1-4.

Coleman, B.L. Salvadori, M.I., McGeer, A.J., Sibley, K.A., Neumann, N.F., Bondy, S.J., Gutmanis, I.A., McEwen, S.A. Lavoie M. Strong, D., Johnson, I., Jamieson, F.B., Louie, M. (2012). The role of drinking water in the transmission of antimicrobial-resistant *E. coli*. Epidemiology and Infection, 140, 633–642.

Committee on Environmental Health and Committee on Infectious Disease. (2009) Drinking water from private wells and risks to children. Pediatrics, 123, 1599-1605.

Corkal, D., Schutzman, W., Hilliard, R. (2004). Rural water safety from the source to the on-farm tap. Journal of Toxicology and Environmental Health, 67, 1619-1642.

Courvoisier, D., Combescure, C., Agoritas, T., Gayet-Ageron, A., Perneger, T. (2011). Performance of logistic regression modeling: beyond the number of events per variable, the role of data structure. Journal of Clinical Epidemiology, 64, 993-1000.

Craun, G., Brunkard, J., Yoder, J., Roberts, V., Carpenter, J., Wade, T., Calderon, R., Roberts, J., Beach, M., Roy, S. (2010). Causes of outbreaks associated with drinking water in the United States from 1971-2006. Clinical Microbiology Reviews, 23, 507-528.

Curriero, F. Patz, J., Rose, J., Lele, S. (2001). The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. American Journal of Public Health, 91, 1194-1199.

Denno, D., Keene, W., Hutter, C., Koepsell, J., Patnode, M., Flodin-Hursh, D., Stewart, L., Duchin, J., Rasmussen, L., jones, R., Tarr, P. (2009). Tri-county comprehensive assessment of risk factors for sporadic reportable bacterial enteric infection in children. The Journal of Infectious Diseases, 199, 467-476.

Elena, S., Whittam, T., Winkworth, C., Riley, M., Lenski, R. (2005). Genomic divergence of *Escherichia coli* strains: Evidence for horizontal transfer and variation in mutation rates. International Microbiology, 8, 271-278.

Environment Canada. (2006). A primer on fresh water: questions and answers. Minister of the Environment. www.ec.gc.ca/water/ Accessed October 24, 2013

Frazier, P., Barron, K., Tix, A. (2004). Testing moderator and mediator effects in counseling psychology. Journal of Counseling Psychology, 51, 115–134.

Frind, E., Molson, J., Rudolph, D. (2006). Ground water: Well vulnerability: A quantitative approach for source water protection. Ground Water, 44, 732–742.

Gallay, A., De Valk , H., Cournot , M., Ladeuil, B., Hemery , C., Castor, C., Bon, F., Megraud, F., Cann , P., Desenclos, J. (2006). A large multi-pathogen waterborne community outbreak linked to faecal contamination of a groundwater system, France, 2000. Clinical Microbiology and Infection, 12, 561–570.

Gannon, V. P. J..Graham, T.A., Read, S., Ziebell, K., Muckle, A., Mori, J., Thoman, J., Selinger, B., Townshend, I., Bryne, J. (2004). Bacterial pathogens in rural water supplies in Southern Alberta, Canada. Journal of Toxicology and Environmental Health, 67, 643-1653.

George, I., Anzil, A., Servais, P. (2004). Quantification of fecal coliform inputs to aquatic systems through soil leaching. Water Research, 38, 611–618.

Gorelick, M., McLellan, S.,Wagner, D., Klein, J. (2011). Water use and acute diarrhoeal illness in children in a United States metropolitan area. Epidemiology and Infection, 139, 295-301.

Goss, M., Barry, D., Rudolph, L. (1998). Contamination in Ontario farmstead domestic wells and its association with agriculture: Results from drinking water wells. Journal of Contaminant Hydrology, 32, 267–293.

Hancock, D., Besser, T., Lejeune, J., Davis, M., Rice, D. (2001). The control of VTEC in the animal reservoir. International Journal of Food Microbiology, 66, 71–78.

Health Canada. (2012). Guidelines for Canadian Drinking Water Quality: Guideline Technical Document. Ottawa, Ontario: Health Canada.

Heinze, G., Schemper, M. (2002). A solution to the problem of separation in logistic regression. Statistics in Medicine, 21, 2409–2419.

Helms, M., Simonsen, J., Olsen, K., Mølbak, K. (2005). Adverse health events associated with antimicrobial drug resistance in *Campylobacter* species: A registry-based cohort study. The Journal of Infectious Diseases, 191, 1050–1055.

Hexemer, A., Pintar, K., Bird, T., Zentner, S., Garcia, H., Pollari, F. (2008). An investigation of bacteriological and chemical water quality and the barriers to private well water sampling in a Southwestern Ontario community. Journal of Water Health, 6, 525-525.

Horswell, J., Hewitt, J., Prosser, J., Van Schaik, A., Croucher, D., Macdonald, C., -Burford, P., Susarla, P., Bickers, P., Speir, T. (2010). Mobility and survival of *Salmonella typhimurium* and human adenovirus from spiked sewage sludge applied to soil columns. Journal of Applied Microbiology, 108, 104–114.

Hosmer, D.W, Lemeshow, S. (2000). Applied logistic regression: Second edition. New York: John Wiley & Sons, INC.

Hunter, P., Colford, J., LeChevallier, M., Binder, S., Berger, P. (2001). Waterborne diseases. Emerging Infectious Diseases, 7, 544-545.

Hussein, H. (2007). Prevalence and pathogenicity of Shiga toxin-producing *Escherichia coli* in beef cattle and their products. Journal of Animal Science, 85, E63-E72.

Isaac-Renton, J., Blatherwick, J., Bowie, W., Fyfe, M., Khan, M., Li, A. (1999). Epidemic and endemic seroprevelance of antibodies to *Cryptosporidium* and *Giardia* in residents of three communities with different drinking water supplies. American Journal of Tropical Medicine and Hygiene, 60, 578-583.

Jackson, S., Goodbrand, R., Johnson, R., Odorico, V., Alves, D., Rahn, K., Wilson, J,, Welch, M., Khakhria, R. (1998). *Escherichia coli* O157:H7 diarrhoea associated with well water and infected cattle on an Ontario farm. Epidemiology and Infection, 120, 17-20.

Jakopanec, I., Borgen, K., Vold, L., Lund, H., Forseth, T., Hannula, R., Nygård, K. (2008). A large waterborne outbreak of campylobacteriosis in Norway: The need to focus on distribution system safety. BMC Infectious Diseases, 8, 128-136.

Jokinen, C., Edge, T., Koning, W., Laing, C., Lapen, D., Miller, J., Mutschall, A., Taboada, E., Thomas, J., Top, E., Wilkes, G., Gannon, V. (2012). Spatial and temporal drivers of zoonotic pathogen contamination of an agricultural watershed. Journal of Environmental Quality, 41, 242–252.

Kapperud, G., Stenwig, H., Lassen, J. (1998). Epidemiology of *Salmonella typhimurium* O:4-12 infection in Norway: evidence of transmission from an avian wildlife reservoir. American Journal of Epidemiology, 147, 774-782.

Karama, M., Johnson, R., holtslander, S., McEwan, S., Gyles, C. (2008). Prevalence and characterization of verotoxin-producing *Escherichia coli* (VTEC) in cattle from an Ontario abattoir. The Canadian Journal of Veterinary Research, 72, 297–302.

Kleinbaum, D., Kupper, L., Nizam, A., Muller, K. (2008) Applied regression analysis and other multivariable methods, 4th Edition. Belmont, California: Brooks/Cole Cengage Learning.

Kreutzwiser, R., de Loe, R., Imgrund, K., Conboy, M., Simpson, H., Plummer, R. (2011). Understanding stewardship behaviour: Factors facilitating and constraining private water well stewardship. Journal of Environmental Management, 92, 1104-1114.

Kreutzwiser, R., de Loe, R., 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. University of Waterloo: Water Policy and Governance Group.

La Ragione, R., Best, A., woodward, M., Wales, A. (2009). *Escherichia coli* O157:H7 colonization in small domestic ruminants. FEMS Microbiology Reviews, 33, 394-410.

Leclerc, H., Schwartzbrod, L., Dei-Cas, E. (2002). Microbial agents associated with waterborne diseases. Critical Reviews in Microbiology, 28, 371–409.

MMWR. (1999) Public health dispatch: Outbreak of *Escherichia coli* O157:H7 and *Campylobacter* among attendees of the Washington County Fair – New York, 1999. MMWR Weekly, 48, 803.

Mac Kenzie, W, Hoxie, N., Proctor, M., Gradus, M., Blair, K., Peterson, D., Kazmierczak, J., Addiss, D., Fox, K., Rose, J., Davis, J. (1994). A massive outbreak in Milwaukee of *cryptosporidium* infection transmitted through the public water supply. The New England Journal of Medicine, 331, 161-167.

MacKinnon, D., Krull, J., Lockwood, C. (2000). Equivalence of the mediation, confounding and suppression effect. Prevention Research, 1, 173-185

Martin, S., Penttinen, P., Hedin, G., Ljungstrom, M., Allestam, G., Andersson, Y., Giesecke, J. (2006). A case–cohort study to investigate concomitant waterborne outbreaks of *Campylobacter* and gastroenteritis in Soderhamn, Sweden, 2002–3. Journal of Water and Health, 4, 417-424.

Mathis, J., Brye, K., Dunn, S. (2011). Preliminary evaluation of septic system absorption field architecture types in a profile-limited soil. Journal of Environmental Quality, 40, 1661–1673.

Michel, P., Wilson, J., Martin, S., Clarke, R., McEwen, S., Gyles, C. (1999). Temporal and geographical distributions of reported cases of *Escherichia coli* O157:H7 infection in Ontario. Epidemiology and Infection, 122, 193-200.

Minnesota Department of Agriculture: Animal Calculation Worksheet**.** *Minnesota Department of Agriculture* 2005. www.mda.state.mn.us Accessed May 2013.

Naylor, S., Gallyb, D., Low, C. (2005). Enterohaemorrhagic *E. coli* in veterinary medicine. International Journal of Medical Microbiology, 295, 419–441.

New York State Department of Health. (2008). Individual water supply wells – Fact sheet #5. Bureau of Water Supply Protection: Troy, New York.

Nygard, K., Anderson, Y., Rottingen, J., Svensson, A., Lindback, J., Kistemann, T. (2004). Association between environmental risk factors and *Campylobacter* infections in Sweden. Epidemiology and Infection, 132, 317-325.

O'Reilly, C., Bowen, A., Perez, N., Sarisky, J., Shepard, C., Miller, M., Hubbard, B., Herring, M., Buchanan, S., Fitzgerald, C., Hill, V., Arrowood M., Xiao, L., Hoekstra, R., Mintz, E., Lynch, M., Outbreak Working Group. (2007). A waterborne outbreak of gastroenteritis with multiple etiologies among resort island visitors and residents: Ohio, 2004, Clinical Infectious Diseases, 44, 506–512.

Olsen,S., Milller, G., Breuer, T., Kennedy, M., Higgins,C., Walford J., McKee, G., Fox, K., Bibb, W., Mead, P. (2002). A waterborne outbreak of *Escherichia coli* 0157:H7 infections and hemolytic uremic syndrome: Implications for rural water systems. Emerging Infectious Diseases, 8, 370-375.

Pebody, R.. Furtado, C., Rojas, A., McCarthy, N., Nylen, G., Ruutu, P., Leino, T., Chalmers, R., DeJong, B., Donnelly, M., Fisher, I., Gilham, C., Graverson, L., Cheasty, T., Willshaw, G., Navarro, M., Salmon, R., Leinikki, P., Wall, P., Bartlett, C. (1999). An international outbreak of vero cytotoxin-producing *Escherichia coli* O157 infection amongst tourists; a challenge for the European infectious disease surveillance network. Epidemiology and Infection, 123, 217-223.

Renter, D., Smith, D., King, R., Stilborn, R. Berg, J., Berezowski, J., McFall, M. (2008). Detection and determinants of *Escherichia coli* O157:H7 in Alberta feedlot pens immediately prior to slaughter. The Canadian Journal of Veterinary Research, 72, 217–227.

Reynolds, K., Mena, K., Gerba, C. (2008). Risk of waterborne illness via drinking water in the United States. Reviews of Environmental Contamination and Toxicology, 192, 117–158.

Richardson, H., Nichols, G., Lane, C., Lake, I., Hunter, P. (2009). Microbiological surveillance of private water supplies in England – The impact of environmental and climate factors on water quality. Water Research, 43, 2159 –2168.

Ritter, L., Solomon, K., Sibley, P., Hall, K., Keen, P., Licence, K., Oates, K., Synge, B., Reid, T. (2001). An Outbreak of *E. coli* 0157 infection with evidence of spread from animals to man through contamination of a private water supply. Epidemiology and Infection, 126, 135-138.

Ritter, L., Solomon K., Sibley, P., Hall, K., Keen, P., Mattu, G., Linton, B. (2002) Sources, pathways and relative risks of contaminants in surface water and groundwater: A perspective prepared for the Walkerton Inquiry. Journal of Toxicology and Environmental Health, 65, 1-142.

Rosenberg, T. (1997). Shigellosis on Indian reserves in Manitoba, Canada: Its relationship to crowded housing, lack of running water and inadequate sewage disposal. American Journal of Public Health, 87, 1547-1551.

Savichtcheva, O., Okabe, S. (2006). Alternative indicators of fecal pollution: relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. Water Research, 40, 2463-2476.

Schuster, C., Ellis, A., Robertson, W., Charron, D., Aramini, J., Marshall, B., Medeiros, D. (2005). Infectious disease outbreaks related to drinking water in Canada, 1974-2001. Canadian Journal of Public Health, 96, 254-258.

Shook, G., Grantham, C. (1993). A decision analysis technique for ranking sources of groundwater pollution. Journal of Environmental Management, 37, 201-206.

Simpson, Hugh. (2004) Promoting the management and protection of private water wells. Journal of Toxicology and Environmental Health, 67, 1679-1704.

Statistics Canada. Households and the environment. 2009. 2011.

Steinberg, K, Levin, B. (2007). Grazing protozoa and the evolution of the *Escherichia coli* O157:H7 Shiga toxin-encoding prophage. Proceedings of the Royal Society of Biological Sciences, 274, 1921-1929.

Stirling, R., Aramini, J., Ellis, A., Lim, G., Meyers, R., Fleury, M., Werker, D. (2001). North Battleford, Saskatchewan, Spring 2001, Waterborne Cryptosporidiosis Outbreak. Guelph, Ontario: Health Canada.

Summers, R. (2010). Alberta Water Well Survey. A report prepared for Alberta Environment. University of Alberta: Edmonton, Canada.

Surampalli, R., Lai, K., Banerji, S., Smith, J., tyagi, R., Lohani, B. (2008). Long-term land application of biosolids–a case study. Water Science & Technology, 57, 345-352.

Swistock, B., Clemens, S., Rummel, S. (2013). Water quality and management of private drinking water wells in Pennsylvania. Journal of Environmental Health, 75, 60-66.

Technical Advisory Committee on Safe Drinking Water. (2004). Environmental public health field manual for private, public and communal drinking water systems in Alberta, Second Edition. Alberta Health: Edmonton, Alberta.

Thomas, K., Charron, D., Waltner-Toews, D., Schuster, C., Maarouf, A., Holt, J. (2006). A role of high impact weather events in waterborne disease outbreaks in Canada, 1975 – 2001. International Journal of Environmental Health Research, 16, 167-180.

Thomas, K., Murray, R., Flockhart, L., Pintar, K., Pollari, F., Fazil, A., Nesbitt, A., Marshall, B. (2013). Estimates of the burden of foodborne illness in Canada for 30 specified pathogens and unspecified agency, circa 2006. Foodborne Pathogens and Disease, 10, 639-648.

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. Journal of Water and Health, 07, 692-698.

Valcour, J., McEwen, M., Wilson, J. (2002). Association between indicators of livestock farming intensity and incidence of human shiga toxin-producing *Escherichia coli* infection. Emerging Infectious Diseases, 8, 252-257.

Williams, B., Healy, D. (2001). Perceptions of illness causation among new referrals to a community mental health team: "explanatory model" or "exploratory map"? Social Science & Medicine, 53, 465-476.

Won, G., Gill, A., LeJeune, J. (2013). Microbial quality and bacteria pathogens in private wells used for drinking water in northeastern Ohio. Journal of Water and Health, 11, 555-562.

Yoon, J., Hovd, C. (2008). All blood, no stool: enterohemorrhagic *Escherichia coli* O157:H7 infection. Journal of Veterinary Science, 9, 219-231.

Yuling, L., Frey, E., Mackenzie, A., Finlay, B. (2000). Human response to *Escherichia coli* O157:H7 infection: Antibodies to secreted virulence factors. Infection and Immunity, 68, 5090-5095.

Ontario Well Water Study

Household Questionnaire

Date of interview: ☐☐ - ☐☐ - ☐☐☐☐ (dd/mm/yyyy)

Interviewer: ____________________________________

Household ID: ☐☐☐☐

☐ **Consent acquired**

I am going to start with a general household questionnaire. It should take about 10 minutes and covers things about people who live here, your water supply and septic system, and even your pets.

You are free to refuse to answer any question and to stop the interview at any time. However, your answers are all important and I hope you are able to answer all of the questions I ask you.

Do you have any questions?

I am going to start with a few questions about you.

1. The respondent is:
   - ☐ Male
   - ☐ Female

2. How old are you?
   ☐☐☐ years *(999 for don't know/refused)*

3. How long have you lived at <u>this</u> address? *(using this well)*
   ☐☐☐ months
   ☐☐☐ years *(999 for don't know/refused)*

I am going to ask about people who currently live in your home. For these questions, I would like to know about people who live in your home, whether or not they are related to you, but who live at this address four or more days per week.

4. Including yourself, how many adults, that is people 20 years and older, currently live in your home?
   ☐☐ (99 don't know/refused)

5. How many youths 12 to 19 live here?

☐☐ *(If none, enter 0, do NOT leave blank)*

6. And how many children 4 to 11 years?
   ☐☐
7. How many children under 4 years of age live in your household?
   ☐☐ (If zero, skip to Q=8)

   7a. Are any of the children still in diapers? *(Includes "pull-ups")*
   - ☐ Yes
   - ☐ No
   - ☐ Don't know / refused

   7b. Do any children in your household go to a day care centre?
   (5 or more children in centre; child in care 1 or more days/week)
   - ☐ Yes
   - ☐ No
   - ☐ Don't know

8. Does anyone in the household work at any of the following…

	Yes	*No*
Day care centre or babysitting service	☐	☐
Hospital, nursing home or residential home	☐	☐
Sewage treatment plant	☐	☐
Any other job where they are in contact with human waste: ______________________	☐	☐
Farm with livestock (any type)	☐	☐
Abattoir, butcher shop, or meat processing	☐	☐
Animal feed processing plant	☐	☐
Nursery or landscaping service	☐	☐
Any other job where they are in contact with meat, animals, or animal waste:___________	☐	☐

9. What township and county is this residence [property] a part of?

Township: ____________________________________ (Write don't know/refused as required)
County: ______________________________________

10. Do you have a swimming pool? (*not a pond or swimming hole*)
    - ☐ Yes
    - ☐ No
    - ☐ Don't know

11. Do you have a hot tub or spa?
    - ☐ No
    - ☐ Yes

12. Have many washrooms do you have in your home?

☐☐ *(With toilet and sink i.e. outhouse=0; 99 for don't know/refused)*

Now I would like to ask a few questions about your pets.

13. Do you have any pets?
    - ☐ Yes
    - ☐ No *(Skip to Q=14)*
    - ☐ Don't know/refused

13a. What kind of pets do you have?
- ☐ Dog(s)
- ☐ Cat(s)
- ☐ Bird(s)
- ☐ Other: specify: ______________________________

14. In the past three months, have any animals spent more than a few minutes <u>inside</u> the house? *(Several hours per week. Include animals that only live in house e.g. hamster)*
    - ☐ Yes
    - ☐ No *(Skip to Q=17)*
    - ☐ Don't know/refused

14a. What kind of animals have spent time inside the house?
- ☐ Dog(s)
- ☐ Cat(s)
- ☐ Bird(s) - *(Skip to Q=16)*
- ☐ Other: specify: ______________________________ - *(Skip to Q=16)*

______________________________ - *(Skip to Q=16)*

15. How often would you say you give your <FILL: cat and/or dog> any of the following. Would you say your pet(s) often, sometimes, rarely or never get(s)… *(Read list)*

	Often	*Some Times*	*Rarely*	*Never*
Commercial dry or canned food	☐	☐	☐	☐
Commercial biscuits or dry treats	☐	☐	☐	☐
Raw meat (any kind)	☐	☐	☐	☐
Cooked meat	☐	☐	☐	☐
Raw hide treats	☐	☐	☐	☐

16. Were <u>any</u> of these animals on antibiotics in the past <u>three months</u>? *(Not just cat/dog. Ref: calendar)*
    - ☐ Yes
    - ☐ No *(Skip to Q=17)*
    - ☐ Don't know/don't remember

16a. Do you recall what kind(s) of antibiotic were they given?

____

____

____

____

I'm going to ask a few questions about your drinking water now.

17. Where do you get the water you use for drinking, food preparation, and dental care? Is it from a private well, a well used by 6 or more households, a cistern, a municipal system, or some other source?

- ☐ Private well
- ☐ Communal well *(6 or more households)*
- ☐ Cistern *(Skip to Q=21)*
- ☐ Municipal (or town) water *(Skip to Q=21)*
- ☐ Other: ____________________________
- ☐ Don't know

18. What type of well do you have? Is it drilled, dug, bored, or driven, which is also called a sand point or well point?

- ☐ Drilled
- ☐ Dug or bored
- ☐ Driven *(sandpoint or wellpoint)*
- ☐ Other: _______________________
- ☐ Don't know

19. How deep is your well?

☐☐☐☐ feet *(9999 for don't know)*

☐☐☐☐ metres

20. How old is it?

☐☐☐ months *(999 for don't know)*

☐☐☐ years

21. Have any repairs or maintenance be done on your well or water lines in the past <u>year</u>?

- ☐ Yes
- ☐ No → skip to Q=22
- ☐ Don't know

21a. And in the past three months have any repairs or maintenance be done on your well or water lines?

- ☐ Yes
- ☐ No
- ☐ Don't know

22. Why did you submit your water for bacteriological testing <u>this most recent time</u>?

- ☐ Do it regularly / routinely
- ☐ Off colour / cloudy
- ☐ Bad / different taste
- ☐ Odour
- ☐ Heavy rain
- ☐ People ill with stomach illness / diarrhoea
- ☐ *E. coli* in previous test
- ☐ Coliforms in previous test
- ☐ Other: ____
- ☐ No specific reason
- ☐ Don't know

23. How many times did you send your water for bacteriological testing in the past
<u>12 months</u>?

- ☐☐ Number
- ☐ Many times *(don't know exactly, but more than 10)*
- ☐ Don't know

23a. On average how many times have you sent your water for testing in the past five years?

- ☐☐ Number
- ☐ Many times *(don't know exactly, but more than 10)*
- ☐ Don't know

24. And do you recall how many times it tested positive for coliforms?

- ☐☐ Number
- ☐ Many times *(don't know exactly but more than 10)*
- ☐ Don't know

24a. How many times has your well water tested <u>positive</u> for <u>*E. coli*</u> in the past
12 months?

- ☐☐ Number
- ☐ Many times *(don't know exactly but more than 10)*
- ☐ Don't know

25. Do you currently <u>treat</u> the water you use for drinking? By treating, I mean boiling, adding chlorine or some other treatment to remove bacteria and other contaminants?

- ☐ Yes
- ☐ No *(Skip to Q=26)*
- ☐ Don't know

25a. How do you treat it?

- ☐ Boil
- ☐ Chlorine → Did you "shock" treat it ☐ or is this an ongoing treatment ☐ ?
- ☐ Filtration
- ☐ Brita or other "filter" system
- ☐ Ultraviolet (UV)
- ☐ Ozone
- ☐ Other: ______________________________
- ☐ Don't know

25b. When did you start treating it ?

☐☐ - ☐☐ - ☐☐☐☐ (dd/mm/yyyy) *(Year only if several years ago)*

25c. And do you treat the water you use for food preparation, dental care, and bathing?

- ☐ Yes
- ☐ No *(Skip to Q=26)*
- ☐ Don't know

25d. How do you treat it?

- ☐ Boil
- ☐ Chlorine → Did you "shock" treat it ☐ or is this an ongoing treatment ☐ ?
- ☐ Filtration
- ☐ Brita or other "filter" system
- ☐ Ultraviolet (UV)
- ☐ Ozone
- ☐ Other: ______________________________
- ☐ Don't know

Now a few questions about your septic system. Remember that everything you tell me during this survey is confidential. Your name will not be connected to anything you tell me and it will never be shared with anyone outside this study.

26. How is your domestic sewage handled? Do you have a … *(Read list)*

- ☐ Septic tank and weeping bed *(aka: field or leaching bed)*
- ☐ Field tank
- ☐ Holding tank
- ☐ Lagoon
- ☐ Surface discharge *(Skip to Q=28)*
- ☐ Municipal system *(Skip to Q=28)*
- ☐ Other: ____________________________________
- ☐ Don't know ***Do NOT read***

27. When was the last time you had the tank [lagoon] pumped?

☐☐☐ *months (999 for don't know)*

☐☐☐ years (888 for never)

28. How old is your septic system? [or – How long have you been on the municipal sewage system?] *(Note: Oldest part if renovations completed)*

☐☐☐ *months (999 for don't know)*
☐☐☐ years

29. Have any upgrades or maintenance been done on your sewage system in the past year?

- ☐ Yes
- ☐ No → skip to Q=30
- ☐ Don't know

29a. And in the past three months have any repairs or maintenance be done on your sewage system??

- ☐ Yes
- ☐ No
- ☐ Don't know

Next I would like to ask a few things about your property.

30. How would you describe the soil on this property. Would you say it is predominantly *(Read list)*

- ☐ Gravel
- ☐ Sand
- ☐ Loam, or
- ☐ Clay
- ☐ Don't know ***Do NOT read***

31. How many acres of property do you own [rent] <u>at this location</u>?

☐☐☐☐ acres (9999 for don't know)
☐☐☐☐ hectares

32. Would you describe your property as being … *(Read list)*

- ☐ Farm
- ☐ Non-farm rural (Skip to Q=33)
- ☐ Village or hamlet (<1,000 people) *(Skip to Q=39)*
- ☐ Small town (1,000 to 10,000 people) *(Skip to Q=39)*
- ☐ Other: ________________________________

32a. What percentage of the land is tilled?

__%

33. Have livestock been housed on this property in the past 12 months? This includes animals owned and/or cared for by your family or housed here and cared for by other people.

- ☐ Yes
- ☐ No *(Skip to Q=39)*
- ☐ Don't know

33a. What type of livestock have been on this property in the past 12 months? *(Check all that apply.)*

- ☐ Dairy cattle
- ☐ Beef cattle
- ☐ Sheep (lambs)
- ☐ Goats
- ☐ Pigs
- ☐ Horses (ponies)
- ☐ Chickens
- ☐ Turkeys
- ☐ Other: specify ______________________________

____

33b. What is the largest number of <FILL: *type of livestock*> that have been housed on this property in the past 12 months?

________________________________ (type) ☐☐☐☐☐☐
________________________________ (type) ☐☐☐☐☐☐
________________________________ (type) ☐☐☐☐☐☐
________________________________ (type) ☐☐☐☐☐☐

34. Are livestock currently housed on the property?

- ☐ Yes
- ☐ No
- ☐ Don't know

35. Do you care for the livestock on this property?

- ☐ Yes
- ☐ No *(Skip to Q=37)*
- ☐ Don't know

36. Have you used antibiotics as a feed supplement for your livestock in the past 12 months?

- ☐ Yes
- ☐ No (*Skip to Q=37)*
- ☐ Don't know

36a. What type(s) of antibiotic have you used as a feed supplement?

____________________________________ *(Write in 'don't know' if applicable)*
____
____

36b. When did you start using antibiotics in your feed?

☐☐ - ☐☐☐☐ *(mm/yyyy)* *(Year is sufficient if several years ago)*

36c. Are you still using antibiotics in your feed?

- ☐ Yes *(Skip to Q=37)*
- ☐ No
- ☐ Don't know

36d. When did you stop using it?

☐☐ - ☐☐☐☐ *(mm/yyyy)*

36e. How do you dispose of unused feed?

- ☐ Spread on fields
- ☐ No feed unused
- ☐ Other: ______________________________
- ☐ Don't know

37. Where, <u>in relation to your well</u>, has manure been <u>stored or spread</u> over past year? Would you say it is stored or spread …*(Read list)* *(Includes liquid or solid; stored in any way: piled, in feed lot, in lagoon, cement or steel tank, etc.)*

- ☐ Within 15 metres (50') of your well
- ☐ Within 30 metres (100') of the well
- ☐ Within 100 metres (330') of your well *(Skip to Q=39)*
- ☐ More than 100 metres (330') from the well *(Skip to Q=39)*
- ☐ Not spread in past 12 months *(Skip to Q=39)*
- ☐ Don't know ***Do NOT read***

37a. When was the <u>last time</u> manure was stored or spread on fields <u>within 30 metres</u> (100') of your well? Would you say… *(Read list)*

- ☐ Within the past month
- ☐ Within the past 3 months
- ☐ Within the past 12 months
- ☐ More than 12 months ago, or
- ☐ Never
- ☐ Don't know ***Do NOT read***

38. How soon is manure <u>usually</u> worked into the ground when it is spread? Would you say it is worked in… *(Read list)*

- ☐ Same day (includes injected)
- ☐ Within 1 to 3 days
- ☐ Within 4 to 7 days
- ☐ More than one week after it is spread
- ☐ Don't know ***Do NOT read***

39. Has a neighbour <u>bordering your property</u> had livestock on their land in the past 12 months? By bordering, I mean a neighbour that shares a fence line with you.

- ☐ No *(Skip to Q=42)*
- ☐ Yes
- ☐ Don't know

39a. What type of livestock were on that property within the past 12 months?

- ☐ Dairy cattle
- ☐ Beef cattle
- ☐ Sheep
- ☐ Goats
- ☐ Pigs
- ☐ Horses
- ☐ Chickens
- ☐ Turkeys
- ☐ Other: specify ______________________________
____
____

40. Has a neighbour bordering on your property spread manure on their fields in the past year?

- ☐ Yes
- ☐ No *(Skip to Q=42)*
- ☐ Don't know

40a. Where, in relation to your well, have neighbours spread manure over past year? Would you say it is spread … *(Read list)*

- ☐ Within 15 metres (50') of well
- ☐ Within 30 metres (100') of well
- ☐ Within 100 metres (330') of well *(Skip to Q=42)*
- ☐ More than 100 metres (330') from well *(Skip to Q=42)*
- ☐ Not spread in past 3 months *(Skip to Q=42)*
- ☐ Don't know ***Do NOT read***

40b. When was the last time manure was spread on fields within 30 metres (100') of your well? Would you say… *(Read list)*

- ☐ Within the past month
- ☐ Within the past 3 months
- ☐ Within the past 12 months ago
- ☐ More than 12 months ago, or
- ☐ Never
- ☐ Don't know ***Do NOT read***

41. How soon is manure <u>usually</u> worked into the ground when it is spread? Would you say… *(Read list)*

- ☐ Same day (includes injected)
- ☐ Within 1 to 3 days
- ☐ Within 4 to 7 days
- ☐ More than one week
- ☐ Don't know ***Do NOT read***

42. Do you fertilize your vegetable or flower gardens or fruit orchards with animal manure? (any source including store purchased)

- ☐ Yes
- ☐ No
- ☐ Don't know

43. Has sludge from <u>human</u> waste been spread on fields within 90 metres (300 feet) of
your well in the past 12 months?

- ☐ Yes
- ☐ No
- ☐ Don't know

44. And in the past 12 months, has waste from meat processing been spread within 90
metres (or 300 feet) of your well?

- ☐ Yes
- ☐ No
- ☐ Don't know

45. Did you have flooding anywhere on your property in 2005?

- ☐ Yes
- ☐ No → skip to Q=46
- ☐ Don't know

45a. Did the flood water cover your well head, that is: the top of the well that is at or above ground level?

- ☐ Yes
- ☐ No
- ☐ Don't know

45b. Was the soil covering your septic tank or weeping tile flooded?

- ☐ Yes
- ☐ No
- ☐ Don't know

46. I am going to ask you to estimate some distances on the property. Please give me your best estimate of the distance between the well head and the closest point of the…

	Distance	M	Km	Ft	Yd	Mile	DK	NA
Septic tank	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Weeping tile	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
House	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Garden (vegetable or flower)	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Manure storage	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Stable or kennel (closest)	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Pasture (past 12 months)	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Field where manure applied	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Tilled fields	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Open water	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Forest/wooded area	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Sanitary land fill site	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Nearest property line	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Neighbour's septic system	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐
Municipal sewage tile	☐☐☐☐	☐	☐	☐	☐	☐	☐	☐

NA = not applicable
DK = don't know

46a. Are any of the locations I just listed uphill from the wellhead? Would you say your wellhead is downhill from the…

	Yes	No	DK	NA
Septic tank	☐	☐	☐	☐
Weeping tile	☐	☐	☐	☐
House	☐	☐	☐	☐
Garden (vegetable or flower)	☐	☐	☐	☐
Manure storage	☐	☐	☐	☐
Stables (or kennels)	☐	☐	☐	☐
Pasture (past 12 months)	☐	☐	☐	☐
Field where manure applied	☐	☐	☐	☐
Tilled fields	☐	☐	☐	☐
Open water	☐	☐	☐	☐
Forest/wooded area	☐	☐	☐	☐
Sanitary land fill site	☐	☐	☐	☐
Nearest property line	☐	☐	☐	☐
Neighbour's septic system	☐	☐	☐	☐
Municipal sewage tile	☐	☐	☐	☐

I am going to ask a few questions that will help us group your information with other households most like your own. Remember that nothing about you, as an individual, will ever be released and you are identified by number in this study.

47. First, what is the <u>highest</u> level of education that has been attained by any adult in the household? Would that be (*read list) …*

- ☐ Less than grade 9
- ☐ Some high school
- ☐ Graduated high school
- ☐ College or trade school
- ☐ University
- ☐ Don't know ****Do NOT read****
- ☐ Not stated ****Do NOT read****

48. What is your best estimate of the <u>total</u> <u>off-farm</u> income, before taxes and deductions, of <u>all</u> household members combined, from all sources, in 2004? Was that total household income… *(Read list. Note: include income from government sources)*

- ☐ No off-farm income
- ☐ Less than $20,000
- ☐ $20,000 to less than $40,000
- ☐ $40,000 to less than $60,000
- ☐ $60,000 to less than $80,000
- ☐ $80,000 or more
- ☐ Don't know ****Do NOT read****
- ☐ Not stated ****Do NOT read****

<u>For farming households only (Q32=farm)…</u>

49. What is your best estimate of the <u>net</u> income from your farm, before taxes, in 2004? Was that net income … *(Read list)*

- ☐ Less than $20,000
- ☐ $20,000 to less than $40,000
- ☐ $40,000 to less than $60,000
- ☐ $60,000 to less than $80,000
- ☐ $80,000 or more
- ☐ Don't know ****Do NOT read****
- ☐ Not stated ****Do NOT read****
- ☐ Not applicable: Not a farming property

Move directly to personal questionnaire.

Appendix Two: Variables derived from Household Questionnaires, Public Health Laboratory test and Environment Canada

Variable	Data Source(s)	Item	Variable derived
E. coli water result **(DV)** Categorical: 1-present 0-absent	Public health laboratory analysis Screening call Household questionnaire	Public health laboratory bacteriological analysis of water Screen: Has your water tested positive for *E. coli* in the past year? H24. How many times has your well water tested positive for *E. coli* in the past 12 months?	0-Water not contaminated: no *E. coli* contamination for one year or longer (households without contamination on current test but contamination within past 12 months were not eligible). 1-Water contaminated: contamination with *E. coli*
Type of septic system **(IV)** Categorical: 1-septic tank & weeping bed 2-field tank 3-holding tank 4-lagoon 5-surface discharge 6-municipal 7-other	Household questionnaire	H26. How is your domestic sewage handled?	As stated. Where more than one type was listed, the type most likely to impact the well water was selected (i.e, when holding tank and surface discharge were listed, surface discharge was selected).
Type of well **(IV)** Categorical: 1-drilled 2-dug or bored 3-driven	Household questionnaire	H18. What type of well do you have?	As stated. Households were not eligible if: -reported using a cistern, surface water or municipal system (i.e., not a 'well') -reported don't know

Variable		Data Source(s)	Item	Variable derived
Manure storage in past 12 months **(IV)** Categorical: 1-within 15m of well 2-within 30m of well 3-within 100m of well 4-more than 100m from well 5-not spread in last 12 months 6-don't know or refused		Household questionnaire	H33. Have livestock been housed on this property in the past 12 months? H37a. Where, in relation to your well, has manure been stored or spread over the past year? H40b. Where, in relation to your well, have neighbours spread or stored manure over past 12 months?	Nearest distance of H37a and H40b. Where response to H33 was "no", response for H37a was automatically categorized as "not spread in last 12 months".
Livestock housed on property in past 12 months **(IV)** Categorical: 1-yes 2-no		Household questionnaire	H33. Have livestock been housed on this property in the past 12 months?	As stated. Where response to H32 was "non-farm", "village or hamlet" or "small town", responses that were missing for H33 were automatically categorized as "no".
Type of livestock housed **(M)**	Cattle Categorical: 0-no 1-yes	Household questionnaire	H33. Have livestock been housed on this property in the past 12 months? H33a. What type of livestock have been on this property in the past 12 months?	Dairy cattle and/or beef cattle (as stated) Where response to H33 was "no", responses that were missing for H33a were automatically categorized as "no".
	Poultry Categorical: 0-no 1-yes	Household questionnaire	H33. Have livestock been housed on this property in the past 12 months? H33a. What type of livestock have been on this property in the past 12 months?	Chickens and/or turkeys (as stated) Where response to H33 was no, responses that were missing for H33a=no.

Variable		Data Source(s)	Item	Variable derived
Type of livestock housed **(M)**	Pigs Categorical: 0-no 1-yes	Household questionnaire	H33. Have livestock been housed on this property in the past 12 months? H33a. What type of livestock have been on this property in the past 12 months?	Pigs (as stated) Where response to H33 was no, responses that were missing for H33a=no.
	Horses Categorical: 0-no 1-yes	Household questionnaire	H33. Have livestock been housed on this property in the past 12 months? H33a. What type of livestock have been on this property in the past 12 months?	Horses (as stated) Where response to H33 was no, responses that were missing for H33a=no.
Property designation **(M)** Categorical: 1-farm 2-non-farm rural 3-village or hamlet (<1,000 people) 4-small town (1,000 to 10,000 people) 5-cottage		Household questionnaire	H32. Would you describe your property as being…	As stated.

Variable	Data Source(s)	Item	Variable derived
Total income **(C)** Categorical: 2-less than $40,000 3-$40,000 to less than $60,00 4-$60,000 to less than $80,000 5-$80,000 or more 6-don't know 7-not stated	Household questionnaire	H46. What is your best estimate of the total off-farm income, before taxes and deductions, of all household members combined, from all sources, in 2005? Was that total household income... H47. What is your best estimate of the net income from your farm, before taxes, in 2005? Was that net income…	As stated by adding response to H46 and H47, using the midpoint of the income range reported for each. If H46 or H47 was "don't know" or "not stated", Total Income was automatically categorized as "don't know" or "not stated". If H46 or H47 was missing, Total Income was automatically categorized as missing.
Province **(C)** Categorical: 0-Alberta 1-Ontario	Public health laboratory results		As stated.
Mode of interview **(C)** Categorical: 0-phone 1-site visit	Database provided		As stated.
Soil type **(M)** Categorical: 1-gravel 2-sand 3-loam 4-clay	Household questionnaire	H30. How would you describe the soil on your property? Would you say it is predominately…	As stated. Where response to H30 was missing, soil type was randomly assigned by the proportions in each province.

Variable	Data Source(s)	Item	Variable derived
Distance between well head and septic tank **(M)** Categorical (in meters): 1- ≤ 20 2- 20.1 to 30 3- 30.1 to 50 4- ≥ 60.1	Household questionnaire	H44a. I am going to ask you to estimate some distances on the property. Please give me your best estimate of the distance between the well head and…	Where the response to H44a was missing, the median value, stratified by province, was used. Categories were created based on literature review and legislation, stated (phone interview) or measured (site visit) distance was used to assign the category.
Distance between well head and weeping tile **(M)** Categorical (in meters): 1 - ≤ 20 2- 20.1 to 35 3- 35.1 to 60 4- ≥ 60.1	Household questionnaire	H44a. I am going to ask you to estimate some distances on the property. Please give me your best estimate of the distance between the well head and…	Where the response to H44a was missing, the median value, stratified by province, was used. As. Categories were created based on literature review and legislation, stated (phone interview) or measured (site visit) distance was used to assign the category.
Age of septic system **(M)** Continuous	Household questionnaire	H28. How old is your septic system – or – How long have you been on the municipal sewage system?	As stated. Where the response to H28 was missing, the average age, stratified by province, was used.

Variable	Data Source(s)	Item	Variable derived
Depth of well **(M)** Categorical (in meters): 1- ≤ 10 2-10.1 to 30 3-30.1 to 40 4- ≥ 40.1	Household questionnaire	H19. How deep is your well?	Where response to H19 was missing, the average depth, stratified by well type and province, was used. Categories were created based on literature review and stated well depth was used to assign the category.
Age of well **(M)** Categorical (in years): 1- ≤ 10 2-10.1 to 25 3-25.1 to 35 4- ≥ 35.1	Household questionnaire	H20. How old is it?	Where response to H20 was missing, the average age, stratified by well type and province, was used. Categories were created based on literature review and stated well age was used to assign the category.
Rainfall on day of water sample collection **(M)** Categorical: 0 – no rainfall 1 – ≥ 0.1mm	Date of collection: Public health laboratory County of residence: Household questionnaire or public health laboratory Rainfall data: Environment Canada accessed from: http://climate.weatheroffice.gc.ca/climateData/canada_e.html	Date recorded as water collection date. H9.What township and county is this residence a part of? Rainfall recorded by Environment Canada identified weather stations.	Rainfall reported from the weather station in the county* identified on the day the water sample is recorded as being collected. *=where county is missing, township was used and where both county and township were missing, health region was used.

Variable	Data Source(s)	Item	Variable derived
Rainfall 2 days prior to water sample collection **(M)** Categorical: 0 – no rainfall 1 – ≥ 0.1mm	Public health laboratory Household questionnaire Environment Canada Rainfall data: Environment Canada accessed from: http://climate.weatheroffice.gc.ca/climateData/canada_e.html	Date recorded as water collection date. H9.What township and county is this residence a part of? Rainfall recorded by Environment Canada identified weather stations.	Rainfall reported from the weather station in the county* identified two days prior to the day the water sample is recorded as being collected. *=where county is missing, township was used and where both county and township were missing, health region was used.
Season water sample taken **(M)** Categorical: 1-Spring 2-Summer 3-Fall 4-Winter	Public health laboratory	Date recorded as water collection date.	Samples submitted between March to May=Spring Samples submitted between June to August=Summer Samples submitted between Sept to November=Fall Samples submitted between December to February=Winter
Animal Density **(M)** Categorical: 1-Low 2-Medium 3-High	Animal Density = Animal Units/km^2 Animal Units: Calculated based on Animal Unit Factors developed by the Minnesota Department of Agriculture: http://www.mda.state.mn.us/animals/feedlots/feedlot-dmt/feedlot-dmt-animal-units.aspx Km^2: Household questionnaire	H31. How acres of property do you own or rent at this location?	animal units = (animal unit factor)(# of animals) $1km^2$=247.11 acres Animal Density=Animal Units/ km^2

Appendix Two: Variables derived from Household Questionnaires, Public Health Laboratory test and Environment Canada

DV – Dependent Variable

IV – Independent Variable

C – Confounder

M – Mediator

Appendix Three: Bivariate associations between contamination of well water with *E. coli* and covariates based on logistic regression.

Variable		Odds Ratio	Std. Err.	P-value	95% CI	Decision
Well depth (m)	min to 10	2.34	0.492	<0.001	1.55-3.54	Collinear with Well Type – Out
	10.1 to 30	1.97	0.373	<0.001	1.36-2.85	
	30.1 to 40	1.17	0.252	0.476	0.76-1.78	
	40.1 to max	Ref	.	.	.	
	Overall Test	.	.	<0.001	.	
Well type	Drilled	ref.	.	.	.	In
	Dug or Bored	2.40	0.441	<0.001	1.68 - 3.45	
	Sandpoint	0.33	0.154	0.017	0.13 - 0.82	
	Overall Test	.	.	<0.001	.	
Age of Well (years)	min to 10	ref.	.	.	.	In
	10.1 to 25	1.19	0.229	0.349	0.82 - 1.74	
	25.1 to 35	2.07	0.419	<0.001	1.40 - 3.08	
	35.1 to max	2.18	0.486	<0.001	1.42 - 3.38	
	Overall Test	.	.	<0.001	.	
Distance between well head and weeping tile (m)	min to 20	ref.	.	.	.	Not significant at a p-value of 0.25 – Out
	20.1 to 35	1.12	0.231	0.567	0.75 - 1.68	
	35.1 to 60	0.90	0.226	0.682	0.55 – 1.47	
	60.1 to max	1.05	0.216	0.806	0.70 - 1.57	
	Overall Test	.	.	0.789	.	

Appendix 3 continued…

Distance between well head and septic tank (m)	min to 20	ref.	.	.	.	Not significant at a p-value of 0.25 – Out
	20.1 to 30	0.79	0.162	0.248	0.53 - 1.18	
	30.1 to 50	0.89	0.168	0.533	0.61 - 1.29	
	50.1 to max	1.06	0.205	0.763	0.73 - 1.55	
	Overall Test	.	.	0.511	.	
Type of septic system	Septic tank and weeping bed	ref.	.	.	.	Not significant at a p-value of 0.25 – Out
	Field tank	2.40	2.692	0.434	0.27- 21.60	
	Holding tank	0.84	0.496	0.769	0.27 - 2.67	
	Lagoon	0.30	0.369	0.327	0.03 - 3.33	
	Surface discharge	1.35	0.818	0.619	0.41 - 4.43	
	Municipal	0.80	0.437	0.684	0.28 - 2.33	
	Other	0.75	0.507	0.671	0.20 - 2.82	
	Overall Test	.	.	0.874	.	
Age of Septic System (years)		1.01	0.006	0.018	1.00 - 1.03	In

Appendix 3 continued…

	Gravel	ref.	.	.	.	
	Sand	0.70	0.166	0.129	0.44 - 1.11	
	Loam	0.78	0.171	0.263	0.51 - 1.20	
	Clay	1.10	0.232	0.649	0.73 - 1.67	
Soil Type	Overall Test	.	.	0.078	.	In
	Farm	1.46	0.245	0.024	1.05-2.03	
	Non-farm	Ref	.	.	.	
	Village or hamlet (< 1000 people)	0.79	0.173	0.277	0.51-1.21	
	Small town (1000 to 10 000 people)	1.05	0.308	0.874	0.59-1.86	
	Cottage	1.32	0.589	0.538	0.55-3.16	
Property Designation	Overall Test	.	.	0.056	.	In
	Within 15m (50') of well	1.34	0.742	0.597	0.45-3.97	Collinear with Livestock Housed on Property in last 12 Months – Out
	Within 30m (100') of well	2.68	1.357	0.052	0.99-7.23	
	Within 100m (330') of well	0.83	0.272	0.574	0.43-1.59	
	More than 100m (330') from well	1.34	0.339	0.248	0.82-2.202	
	Not spread in past 12 months	Ref	.	.	.	
	Don't know or refused	5.36	0.31	0.007	1.60-17.99	
Manure storage in past 12 months	Overall Test	.	.	0.005	.	

Appendix 3 continued…

Sludge from human waste spread within 90m of well in past 12 months	Yes	1.01	0.53	0.983	0.36-2.80	Not significant at a p-value of 0.25 – Out
	No	Ref	.	.	.	
Livestock housed on property in last 12 months	Yes	1.76	0.279	0	1.29-2.40	In
	No	Ref	.	.	.	
Livestock housed = Cattle	No	ref.	.	.	.	All Out - perfect prediction with Livestock housed on property
	Yes	1.76	0.347	0.004	1.20 - 2.59	
Livestock housed = Poultry	No	ref.	.	.	.	
	Yes	1.88	0.576	0.04	1.02 - 3.42	
Livestock housed = Pigs*	No	Ref	.	.	.	
	Yes	14.86	17	0.001	2.53 - inf.	
Livestock housed = Horses	No	Ref	.	.	.	
	Yes	1.84	0.409	0.006	1.19 - 2.85	
Animal Density (Animal Units/km)	Low	Ref	.	.	.	Collinear with Livestock Housed on Property in last 12 Months – Out
	Medium	2.03	0.461	0.002	1.30 - 3.17	
	High	1.52	0.309	0.042	1.02 - 2.26	
	Overall Test	.	.	0.002		

Appendix 3 continued…

Season water sample taken	Spring (March to May)	Ref	.	.	.	In
	Summer (June to August)	1.43	0.291	0.075	0.96 - 2.14	
	Fall (September to November)	1.19	0.295	0.477	0.73 - 1.94	
	Winter (December to February)	1.99	0.679	0.042	1.03 - 3.89	
	Overall Test	.	.	0.130	.	
Rainfall Day of sample (mm)	0	Ref	.	.	.	Not significant at a p-value of 0.25 – Out
	>0	1.04	0.155	0.794	0.78 - 1.39	
Rainfall 1 day prior to sample (mm)	0	Ref	.	.	.	
	>0	1.06	0.156	0.685	0.80-1.42	
Rainfall 2 days prior to sample (mm)	0	Ref	.	.	.	
	>0	0.97	0.14	0.860	0.74-1.29	
Rainfall 3 days prior to sample (mm)	0	Ref	.	.	.	
	>0	0.89	0.128	0.406	0.67 - 1.18	
Total Rainfall day of and the 3 days prior to water sample collection (mm)	min to 4	ref.	.	.	.	In
	4.1 to 8	1.22	0.27	0.369	0.79-1.87	
	8.1 to 12	1.31	0.31	0.251	0.83-2.08	
	12.1 to max	1.19	0.21	0.304	0.85-1.68	
	Overall	.	.	0.531	.	

Appendix 3 continued…

Total Household Income	Less than $20 000 to less than $40 000	Ref	.	.	.	
	$40 000 to less than $60 000	0.58	0.179	0.079	0.32-1.06	
	$60 000 to less than $80 000	0.68	0.208	0.213	0.37-1.24	
	$80 000 or more	0.52	0.131	0.009	0.31-0.85	
	Respondent did not know or refused to answer	0.60	0.151	0.042	0.36-0.98	
	Overall Test	.	.	0.113	.	In
Province	Alberta	Ref				Design Variable – In
	Ontario	1.36	0.211	0.046	1.01 - 1.84	
Mode of interview	Compustat	Ref				Design Variable – In
	Site	1.12	0.222	0.581	0.76 - 1.65	

*** -** Exact logistic regression used to calculated OR