Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Ellyn Marder April 13, 2012

Risk Factors for Cryptosporidiosis and Giardiasis Infections in the United States: A Case-Case Comparison in 16 States

By

Ellyn Marder
Master of Public Health

Epidemiology

Patrick Sullivan, PhD, DVM
Committee Chair

Jonathan Yoder, MSW, MPH
Committee Member

Risk Factors for Cryptosporidiosis and Giardiasis Infections in the United States: A Case-Case Comparison in 16 States

By

Ellyn Marder

Bachelor of Science
Florida State University
2010

Thesis Committee Chair: Patrick Sullivan, PhD, DVM

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2012

Abstract

Risk Factors for Cryptosporidiosis and Giardiasis Infections in the United States: A Case-Case Comparison in 16 States
By Ellyn Marder

Cryptosporidiosis and giardiasis are common waterborne enteric pathogens. Typical case-control studies using surveillance data present challenges for identifying risk factors for enteric illness because < 5% of cases are captured. To combat these challenges, case-case comparisons have been suggested and used where cases of other enteric illnesses are used as controls. This study uses cases of cryptosporidiosis, giardiasis, and salmonellosis cases reported to the National Electronic Disease Surveillance System (NEDSS) to identify risk factors for cryptosporidiosis and giardiasis. A total of 9,266 cryptosporidiosis, 14,806 giardiasis, and 64,071 salmonellosis cases reported to NEDSS between 2003 and 2010 from 16 states were included in the analysis. Cryptosporidiosis infection was associated with livestock contact, day care association, drinking untreated water, having a private well as a water source, knowing another ill person, recreational water exposure and travel when compared to salmonellosis infection. Giardiasis infection was associated with day care association, drinking untreated water, knowing other ill persons, recreational water exposures, and travel when compared to salmonellosis infection. This study determined that case-case comparisons can be a useful and efficient tool in outbreak investigations that could lead to more rapid identification of causes and timely implementation of control measures.

Risk Factors for Cryptosporidiosis and Giardiasis Infections in the United States: A Case-Case Comparison in 16 States

By

Ellyn Marder

Bachelor of Science
Florida State University
2010

Thesis Committee Chair: Patrick Sullivan, PhD, DVM

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2012

Table of Contents

Chapter I: Literature Review

Cryptosporidium and *Giardia* are globally distributed obligate parasites that infect the intestinal tract of both humans and animals [1, 2]. The incidence of chlorine-resistance cryptosporidiosis is rising, while incidence of giardiasis has failed to decline significantly despite concerted efforts to reduce transmission via community water systems [3, 4]. Both diseases are under reported, although they are both nationally notifiable [3, 5].

It is estimated that an average of 748,123 cases of cryptosporidiosis occur in the U.S. each year [5]. In 2008, there were 10,500 reported cases of cryptosporidiosis, which was a 62.1% increase in reported cases since 2006. This increase, influence by large outbreaks, may also be due in part to a rise in the number of cases [4]. Although there is a proportion of cases likely associated with undetected outbreaks, it is estimated the < 10% of cryptosporidiosis cases are outbreak-associated [6]. The first drinking water-associated outbreak of cryptosporidiosis occurred in the U.S. in 1984 and the first recreational water-associated outbreak occurred in 1988. Since then, cryptosporidiosis has emerged to become the most recognized cause of recreational water-associated outbreaks in the U.S., with 80 outbreaks reported from 2006 to 2008 [4].

An average of 1,221,564 cases of giardiasis are estimated to occur in the U.S. each year, and 19,140 cases were reported in 2008 [3, 5]. This was a decrease since 2006 [3]. Both cryptosporidiosis and giardiasis infections are under reported, as evidenced by the difference in the number of cases reported in 2008 and the estimated annual burden. Some infections are asymptomatic and, of those that are symptomatic, not all seek medical attention [3, 5]. Furthermore, physicians may not request specific laboratory

testing [4]. However, in 2005, the first treatment of cryptosporidiosis was approved by the FDA, which may have increased testing and therefore reported cases. There are several effective therapies available to treat giardiasis in symptomatic patients [3]. In the U.S., reporting as a nationally notifiable disease began in 1994 and 2002 for cryptosporidiosis and giardiasis, respectively [3, 4]. Giardiasis is not reportable in five states: Indiana, Kentucky, Mississippi, North Carolina, and Texas [3].

There are many species of *Cryptosporidium* that can infect humans. The two most commonly reported species of *Cryptosporidium* causing human illness are *C. parvum* and *C. hominis* [2].Only one species of *Giardia* is found in humans, *G. duodenalis* (also known as *G. lamblia* or *G. intestinalis*). Both *Giardia* and *Cryptosporidium* are transmitted via the fecal-oral route, and infection occurs from ingestion of *Giardia* cysts or *Cryptosporidium* oocysts in contaminated food or water. Cryptosporidiosis and giardiasis infections can cause watery diarrhea, abdominal cramps, bloating, fever, nausea, and weight loss. However, asymptomatic infections are common for both conditions. Transmission may also occur via person-to-person or animal-to-person contact [1, 2]. Upon excretion in feces, both *Giardia* cysts and *Cryptosporidium* oocysts are infectious immediately [7, 8]. It has been reported that ingestion of as few as 10 cysts or oocysts can cause illness and infected persons can shed between 10^8 and 10^9 cysts or oocysts in their stool per day [8-10]. Additionally, shedding can persist for months in the case of giardiasis infection and up to 50 days in cases of cryptosporidiosis infection [8].

Risk factors for cryptosporidiosis and giardiasis infections have previously been identified for outbreak-associated and sporadic illness. Outbreaks and sporadic cases of cryptosporidiosis have been associated with drinking water, recreational water,

unpasteurized milk, contact with farm animals, especially cattle, international travel, and contact with ill persons [6, 11]. A 2004 case-control study aimed to identify risk factors for sporadic cryptosporidiosis infection among immunocompetent patients. International travel was determined to be the most significantly associated risk factor for illness. The study also identified contact with persons under 11 years old with diarrhea, contact with cattle, and recreational water exposure as risk factors [11]. Similarly, a case-control study in England and Wales of data from a national surveillance program reported from 2001 to 2002 identified travel abroad, contact with a patient, and contact with cattle as risk factors for sporadic cryptosporidiosis infections [6]. Risk factors for giardiasis include travel to disease-endemic areas, contact with child care settings, contact with infected persons, drinking water, recreational water, and contact with animals [3]. Furthermore, the age distributions of those with giardiasis and cryptosporidiosis are typically bimodal with one peak occurring in children under the age of 9 years and another in adults aged 35-44 years, supporting reports of person-to-person transmission from children in day care to their care givers. While both cryptosporidiosis and giardiasis display seasonality, it is much more pronounced for cryptosporidiosis, with nearly a tenfold increase in reported cases occurring between June and October when compared to January through March, which coincides with increased outdoor activities. [3, 12].

Interrupting transmission of *Giardia* and *Cryptosporidium* is challenging for the following reasons: first, both parasites have a low infectious dose and are infectious immediately upon being excreted in feces [7]. Second, both *Cryptosporidium* and *Giardia* can infect animals and have zoonotic transmission potential [3, 7]. Lastly,

Cryptosporidium is also relatively small in size and highly resistance to chlorine, making decontamination in water challenging [7].

There are challenges when using enteric disease surveillance data due to < 5% of cases seeking medical care, being tested, diagnosed, and then reported to public health authorities. The characteristics of reported cases likely differ from a randomly selected control population [3-5]. Due to this inefficiency, there is some concern about employing typical case-control methodology to obtain unbiased exposure data from controls. McCarthy and Giesecke recommend that when using a secondary study base for cases, one should carefully consider an appropriate control selection, due to the likely poor representation of the study base in the overall population. Controls need to be selected in a manner that will not result in gathering exposure data in a different way from cases and controls [13]. To address this limitation of case-control studies, some researchers have used reported cases of other enteric diseases as control groups [14-16].

Two studies using this case-case comparison methodology used different types of the same illness and a third compared a group of enteric illnesses to the condition of interest [14-16]. By comparing *Salmonella* Enteritidis to other *Salmonella* serotypes, the researchers were able to have equal sensitivities of case ascertainment, since only a small proportion of *Salmonella* illnesses are reported, as well as reduce recall bias between cases and controls [16]. However, there is one important limitation of this method, which must be considered when selecting controls: if case and control conditions share a causal exposure, the result will be bias toward the null value [13, 15]. A case-case comparison study done in Canada took this limitation into consider in its aim to determine risk factors of sporadic cryptosporidiosis infection using national surveillance data. The controls in

this study were those who reported data of onset with at least one of nine other enteric illnesses during the study period. These illnesses were amebiasis, campylobacteriosis, cyclosporiasis, giardiasis, listeriosis, salmonellosis, shigellosis, verotoxigencic *E. coli*, and yersiniosis. Despite overlap of common exposures, especially with giardiasis, the researchers ultimately decided to retain all nine conditions as controls in order to dilute the baising effects of including a specific diagnostic group that is related to the exposure [14].

Cryptosporidiosis and giardiasis are nationally notifible diseases, meaning that regular information regarding cases is considered necessary for prevention and control of the illness. The list of nationally notifiable conditions is revised periodically through the collaboration of public health officials at state health departments, the Council of State and Territorial Epidemiologists (CSTE), and CDC. Nationally notifiable conditions are included in the annual Nationally Notifiable Disease Surveillance System (NNDSS), which is comprised of data reported through various surveillance systems, such as The National Electronic Disease Surveillance System (NEDSS) [17]. NEDSS began in 2001 and is an initiative developed by CDC to advance the development of efficient, integrated, and interoperable surveillance systems at the state and local levels. While other surveillance systems were designed to address a single condition, NEDSS was designed to interface with relevant data sources. In addition to cryptosporidiosis and giardiasis, other enteric illnesses including amebiasis, campylobacteriosis, cholera, cyclosporiasis, *E. coli*, listeriosis, norovirus, rotovirus, salmonellosis, and shigellosis are reported to NEDSS. [18]. Cases can be reported by state health departments and state

public health laboratories. In addition, state and local health departments investigate cases and include exposure data in their reporting.

While salmonellosis is one of the most commonly reported enteric illnesses in the U.S., there is little overlap between reported risk factors for salmonellosis and those for cryptosporidiosis and giardiasis, therefore would serve well as a control condition in a case-case comparison [19]. From 1987 to 1997, the most commonly reported *Salmonella* serotypes were Typhimurium, Enteritidis, and Heidelberg, accounting for over 50% of the human isolates. During the same time period over 50% of the outbreaks reported were caused by *Salmonella* Enteritidis (46%). *Salmonella* Enteritidis infections have been epidemiologically linked to consumption of chicken and eggs as well as reptile/amphibian contact [20-22]. Various other serotypes, including Javiana, have also been linked to reptile/amphibian contact, and foodborne illness, with 94% of non-typhi *Salmonella* attributable to foodborne transmission [5, 23]. Additional general risk factors for salmonellosis infection have been identified through outbreak investigations and case-control studies using surveillance data. These studies have identified a variety of risk factors including consumption of food products (namely chicken and eggs), contact with amphibians and reptiles, having a private well as a home water source, recreational water exposure, day care attendance, and international travel [20-26].

In an outbreak setting, it is time and resource intensive to assemble a control group. Therefore case-case comparisons could be helpful for outbreak response by reducing the time and resources needed to identify the source of the outbreak. NEDSS is a compilation of multi-year data on cryptosporidiosis, giardiasis, and other enteric illnesses. Since more information is needed to fully understand the risk factors for

sporadic cryptosporidiosis and giardiasis cases, this study aimed to explore the utility of case-case comparisons using NEDSS data and to investigate cryptosporidiosis and giardiasis risk factors.

Chapter II: Manuscript

Risk Factors for Cryptosporidiosis and Giardiasis Infections in the United States: A Case-Case Comparison in 16 States

Abstract

Cryptosporidiosis and giardiasis are common waterborne enteric pathogens. Typical case-control studies using surveillance data present challenges for identifying risk factors for enteric illness because < 5% of cases are captured. To combat these challenges, case-case comparisons have been suggested and used where cases of other enteric illnesses are used as controls. This study uses cases of cryptosporidiosis, giardiasis, and salmonellosis cases reported to the National Electronic Disease Surveillance System (NEDSS) to identify risk factors for cryptosporidiosis and giardiasis. A total of 9,266 cryptosporidiosis, 14,806 giardiasis, and 64,071 salmonellosis cases reported to NEDSS between 2003 and 2010 from 16 states were included in the analysis. Cryptosporidiosis infection was associated with livestock contact, day care association, drinking untreated water, having a private well as a water source, knowing another ill person, recreational water exposure and travel when compared to salmonellosis infection. Giardiasis infection was associated with day care association, drinking untreated water, knowing other ill persons, recreational water exposures, and travel when compared to salmonellosis infection. This study determined that case-case comparisons can be a useful and efficient tool in outbreak investigations that could lead to more rapid identification of causes and timely implementation of control measures.

Introduction

Giardiasis and cryptosporidiosis are the most common intestinal parasites of humans in the United States, with an estimated average of 1,221,564 and 748, 123 cases each year, respectively [5]. In 2008, 19,140 cases of giardiasis were reported to the Centers for Disease Control and Prevention (CDC), which was a decrease from the number of reported cases in 2006 [3]. However, there were 10,500 cases of cryptosporidiosis reported in 2008 which was a 62.1% increase from 2006 [4]. The first U.S. drinking water-associated outbreak of cryptosporidiosis occurred in 1984 and the first U.S. recreational water-associated outbreak occurred in 1988, and cryptosporidiosis has emerged as the most recognized case of recreational water-associated outbreaks in the U.S. Between 2006 and 2008, 80 cryptosporidiosis outbreaks were reported to the CDC [4]. However, it is estimated that < 10% of cryptosporidiosis cases are outbreak-associated [5].

Cryptosporidium and *Giardia* are globally distributed obligate parasites that infect the intestinal tract of both humans and animals. Both are transmitted via the fecal-oral route, and infection occurs from ingestion of as few as 10 *Giardia* cysts or *Cryptosporidium* oocysts in contaminated food or water. Transmission may also occur via person-to-person or animal-to-person contact. Cryptosporidiosis or giardiasis infections can cause watery diarrhea, abdominal cramps, bloating, fever, nausea, and weight loss [1, 2]. Infected persons can shed between 10^8 and 10^9 infectious cysts or oocysts in their stool per day [8-10]. Shedding can persist for months in the case of giardiasis infection and up to 50 days in cases of cryptosporidiosis infection [8].

Risk factors for cryptosporidiosis and giardiasis infections have previously been identified for outbreak-associated and sporadic illness. Risk factors for cryptosporidiosis and giardiasis infections have previously been identified for outbreak-associated and sporadic illness. Outbreaks and sporadic cases of cryptosporidiosis have been associated with drinking water, recreational water, unpasteurized milk, contact with farm animals, especially cattle, international travel, and contact with ill persons [6, 11]. A 2004 case-control study aimed to identify risk factors for sporadic cryptosporidiosis infection among immunocompetent patients. International travel was determined to be the most significantly associated risk factor for illness. The study also identified contact with persons under 11 years old with diarrhea, contact with cattle, and recreational water exposure as risk factors [11]. Similarly, a case-control study in England and Wales of data from a national surveillance program reported from 2001 to 2002 identified travel abroad, contact with a patient, and contact with cattle as risk factors for sporadic cryptosporidiosis infections [6]. In addition to travel, contact with patients, and contact with animals, giardiasis infections have been attributed to contact with child care facilities, drinking water and recreational water [3]. Both cryptosporidiosis and giardiasis typically exhibit a bimodal age distribution with peaks occurring in children under 9 years of age and in adults aged 35-44 years, supporting reports of person-to-person transmission from children in day care to their care givers. The seasonality of cryptosporidiosis infections is much more pronounced than that of giardiasis infections, with nearly a tenfold increase in reported cases occurring between June and October compared to January through March [3, 12].

In the U.S., reporting as a nationally notifiable disease began in 1994 and 2002 for cryptosporidiosis and giardiasis, respectively [3, 4]. Cryptosporidiosis is reportable in all states; however, giardiasis is not reportable in five states: Indiana, Kentucky, Mississippi, North Carolina, and Texas [3]. Both cryptosporidiosis and giardiasis infections are under reported, as evidenced by the difference between the number of reported and estimated number of cases, as are other enteric illnesses. Less than 5% of cases seek medical care, are tested, diagnosed, and reported to public health authorities, therefore the characteristics of reported cases likely differ from a randomly selected control population [3-5]. McCarthy and Giesecke recommend that when using a secondary base for cases, researchers should carefully consider an appropriate control selection, since this distinctive study base is not likely to be well represented in the overall population [13]. To address this limitation, some researchers have used reported cases of other enteric diseases as control groups. Two of these case-case comparison studies used different types of the same illness and a third compared a group of other illnesses to the condition of interest [14-16].

This study aimed to explore the utility of case-case comparisons using surveillance data and to investigate cryptosporidiosis and giardiasis risk factors.

Methods

This analysis compared laboratory-confirmed cases of cryptosporidiosis, giardiasis, and salmonellosis reported to the National Electronic Disease Surveillance System (NEDSS)-Base system. The NEDSS-Base system has been described previously

[18]. Sixteen states (Alabama, Arkansas, Idaho, Maine, Maryland, Montana, Nebraska, Nevada, New Mexico, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, and Wyoming) that used the NEDSS-Base system to report cases of nationally notifiable disease during 2003-2010 were included in this analysis (the remained of states use other systems to report cases). Two states were excluded due to insufficient data and/or reporting time.

There is little overlap of known risk factors between cryptosporidiosis or giardiasis and salmonellosis, therefore salmonellosis serves as the control condition in this study. For enteric illnesses, including cryptosporidiosis, giardiasis, and salmonellosis, additional information for each case about exposure to possible risk factors (e.g. contact with livestock, or consumption of well water) is transmitted via the NEDSS-Base Foodborne Program Area Module (PAM). For this analysis, cryptosporidiosis, giardiasis, and salmonellosis cases reported after a state began transmission to the NEDSS-Base system (known as the "go live date") and before December 31, 2010 were included. In 2010, the population of the NEDSS-Base states comprised 22.7% of the U.S. population. Because giardiasis is not a reportable condition in Texas, the 240 giardiasis cases reported by Texas were excluded from the analysis.

Cases of cryptosporidiosis, giardiasis, and salmonellosis are reported to NEDSS by states based on nationally accepted CSTE case definitions [27]. Cases that met the case definition are the time of transmission were included. Although changes in case definition could have occurred, no attempt was made to standardize the case definitions over the study period. Standard exposure data reported in the Foodborne PAM is collected by each state in varying way (i.e. each state may ask cases about exposures

using different questions and procedures for interviewing cases could differ from state-to-state). For this analysis, it was assumed that cases of cryptosporidiosis and giardiasis were asked about exposures in the 14 days prior to illness onset, while cases of salmonellosis were asked about exposures in the 7 days prior to illness onset. Due to different time periods for finalizing data, the case counts in this study might differ slightly from the numbers reported in CDC's annual Summary of Notifiable Diseases.

For this case-case comparison, cases of cryptosporidiosis and giardiasis were compared with cases of salmonellosis. Odds ratios were adjusted for site, sex, race, and age were obtained. Analyses were performed using SAS 9.2.

Results

A total of 9,266 cryptosporidiosis and 14,806 giardiasis and 64,071 salmonellosis cases were reported to NEDSS during the study period. Texas contributed one-third of all cases (29,062, 33.0%). Nearly one-third of giardiasis cases were reported by Virginia (3,174, 21.4%) and Tennessee (1,505, 10.2%) (Table 1).

Compared with salmonellosis cases, cryptosporidiosis and giardiasis cases were significantly less likely to be hospitalized (salmonellosis, 29.9%; cryptosporidiosis, 18.4%); giardiasis, 9.1%) and giardiasis cases were significantly more likely to be imported from outside the country (salmonellosis, 3.5%; giardiasis, 35.0%) (Table 2). Cryptosporidiosis and giardiasis cases were also less likely to be Hispanic (salmonellosis, 25.3%; cryptosporidiosis, 18.7%; giardiasis, 8.0%). Less than 75% of giardiasis cases were white, compared with 87.8% of cryptosporidiosis and 83.0% of salmonellosis cases.

Giardiasis cases were also less likely to be female compared to salmonellosis cases (46.1% vs. 51.6% respectively).

The highest proportion of cryptosporidiosis, giardiasis, and salmonellosis cases were aged 0 to 9 years (36.9%, 12.9%, 41.2%, respectively) (Figure 1). Both cryptosporidiosis and giardiasis displayed bimodal age distributions with additional peaks in adults aged 30 to 44 years, with 12.9% of cryptosporidiosis cases between the ages of 30 and 40 years. Giardiasis demonstrated a broader secondary peak with 23.7% of cases between the ages of 30 and 44 years.

Cryptosporidiosis cases demonstrated a marked seasonality with over 30% of cases occurring in August alone and over 50% of cases occurring during the late summer months (July – September) (Figure 2). Giardiasis cases also exhibited seasonal trends, although they were less pronounced than cryptosporidiosis. The largest proportion of giardiasis cases (37.5%) also occurred in the late summer months.

In multivariable analysis, after adjusted for site, sex, race, and age, cryptosporidiosis infection was significantly associated with livestock contact, day care association, drinking untreated water, having a private well as a water source, knowing another ill person, recreational water exposure, and travel when compared to salmonellosis infection (Table 3). For cryptosporidiosis infections, being a day care worker was more strongly associated with illness (OR 3.56, 95% CI 2.44 – 5.17) than attending day care (OR 1.13, 95% CI 1.00 – 1.28). Cryptosporidiosis was more strongly associated with treated recreational water exposure (OR 8.80, 95% CI 7.96 – 9.73) than untreated recreational water exposure (OR 4.77, 95% CI 4.25 – 5.37).

Giardiasis was significantly associated with day care association, drinking untreated water, knowing other ill persons, recreational water exposures, and travel when compared to salmonellosis infections and adjusted for site, sex, race, and age (Table 4). Similar to cryptosporidiosis, being a day care worker was more strongly associated with illness (OR 2.73, 95% CI 1.79 – 4.17) than attending day care (OR 1.07, 95% CI 0.92 – 1.25). Unlike cryptosporidiosis, untreated recreational water exposure (OR 3.74, 95% CI 3.32 – 4.22) was strongly associated with giardiasis infections than treated recreational water exposure (OR 2.21, 95% CI 1.93 – 2.52) when compared to salmonellosis.

Several exposures had protective odds ratios for cryptosporidiosis or giardiasis including reptile/amphibian contact, poultry contact, having no water source, and having a new pet.

Discussion

This study used national enteric disease surveillance data and a case-case study design to identify risk factors for cryptosporidiosis and giardiasis infections. The most significant risk factor for cryptosporidiosis infection was treated recreational water exposure, while giardiasis was closely associated with drinking untreated water. A case-case design could enable timely analyses of suspected outbreaks and expand the use of previously collected surveillance data. Use of these data might lead to more rapid identification of outbreak causes and timely implementation of control measures. Expanded use of these surveillance data would also provide evidence for the attribution of these diseases to specific exposures, informing science-based prevention efforts.

Cryptosporidiosis and giardiasis are well-known waterborne pathogens, so it follows that the most significant risk factors identified in this study from 2003 to 2010 were water-related. The most significant risk factor for cryptosporidiosis was treated recreational water exposure, which is expected given the parasite's high chlorine tolerance [7]. The associated with working in childcare settings has not been previously established and warrants further investigation. Giardiasis, however, was most associated with drinking untreated water. Both of these infections also demonstrated expected seasonal trends with the peak of cases occurring during the summer months [3, 12, 28, 29]. The increase in cases during the summer corresponds with the increase of outdoor activities, including recreational water exposure, which contributes to these seasonal trends [3, 4].

Day care association and contact with other ill persons were significant risk factors for both illnesses, a finding observed in previous studies [3-5, 28]. Additionally, livestock contact was identified as a risk factor for cryptosporidiosis but not giardiasis infections. Travel was identified as a risk factor for both infections, despite being known as a risk factor for salmonellosis infections [22, 25].

This study is subject to multiple limitations. First, no lab data were available, which meant that comparisons across *Cryptosporidium* and *Salmonella* species were not possible. Second, number of cases reported in this analysis might not be directly comparable with published NNDSS estimates, because of updates made to case records after publication. Third, states reported to NEDSS over varying lengths of time and case definitions over the study period. Procedures for following up with cases also varied from state to state. Some states may have interviewed all reported cases, while others may only

investigate select cases. States also used different questionnaires to assess exposure to risk factors, which could contribute to misclassification bias. It is unknown whether individual states omitted exposure questions if they were felt to be not relevant to the pathogen of interest, or if a standard set of exposure questions was asked for all enteric illnesses. The high proportion of missing data for some risk factors is likely due to variation in exposure question by pathogen. Finally, risk factors closely associated with both cryptosporidiosis or giardiasis and salmonellosis would not be detected using a case-case study design [13, 15].

Previous studies suing this case-case comparison method have found results generally consistent with typical case-control studies that use healthy controls [16]. These studies also emphasize the importance of collecting standard exposure information for all cases of enteric illness so that they may be included in similar analyses in the future [14]. These researchers have found case-case comparisons to be an efficient method of determining accurate risk factors for enteric illnesses [14-16].

As this study demonstrates, a case-case study design and surveillance data can be a cost- and resource-efficient method for states to investigate outbreaks. By using available surveillance data as a comparison, states can decrease the time to identifying implicated sources in outbreak settings. These data can also be used to investigate the proportional contribution of certain exposures (e.g. food, water, animals) to the burden of enteric disease. However, standard guidelines are needed for data collection and management in order to improve the utility of this method.

References

1. R.C.Andrew, T., *Giardiasis as a re-emerging infectious disease and its zoonotic potential.* International Journal for Parasitology, 2000. **30**(12–13): p. 1259-1267.
2. Chen, X.M., et al., *Cryptosporidiosis.* The New England Journal of Medicine, 2002. **346**(22): p. 1723-31.
3. Yoder, J.S., C. Harral, and M.J. Beach, *Giardiasis surveillance - United States, 2006-2008.* MMWR. Surveillance Summaries : Morbidity and Mortality Weekly Report. Surveillance summaries / CDC, 2010. **59**(6): p. 15-25.
4. Yoder, J.S., C. Harral, and M.J. Beach, *Cryptosporidiosis surveillance - United States, 2006-2008.* MMWR. Surveillance Summaries : Morbidity and Mortality Weekly Report. Surveillance summaries / CDC, 2010. **59**(6): p. 1-14.
5. Scallan, E., et al., *Foodborne illness acquired in the United States--major pathogens.* Emerging Infectious Diseases, 2011. **17**(1): p. 7-15.
6. Hunter, P.R., et al., *Sporadic cryptosporidiosis case-control study with genotyping.* Emerging Infectious Diseases, 2004. **10**(7): p. 1241-9.
7. Dillingham, R.A., A.A. Lima, and R.L. Guerrant, *Cryptosporidiosis: epidemiology and impact.* Microbes and Infection / Institut Pasteur, 2002. **4**(10): p. 1059-66.
8. Cacciò, S.M., et al., *Unravelling Cryptosporidium and Giardia epidemiology.* Trends in Parasitology, 2005. **21**(9): p. 430-437.
9. Chappell, C.L., et al., *Cryptosporidium parvum: intensity of infection and oocyst excretion patterns in healthy volunteers.* The Journal of Infectious Diseases, 1996. **173**(1): p. 232-6.

10. Hellard, M.E., et al., *Prevalence of enteric pathogens among community based asymptomatic individuals.* Journal of Gastroenterology and Hepatology, 2000. **15**(3): p. 290-3.

11. Roy, S.L., et al., *Risk factors for sporadic cryptosporidiosis among immunocompetent persons in the United States from 1999 to 2001.* Journal of Clinical Microbiology, 2004. **42**(7): p. 2944-51.

12. Yoder, J.S. and M.J. Beach, *Cryptosporidiosis surveillance--United States, 2003-2005.* MMWR. Surveillance Summaries : Morbidity and Mortality Weekly Report. Surveillance summaries / CDC, 2007. **56**(7): p. 1-10.

13. McCarthy, N. and J. Giesecke, *Case-case comparisons to study causation of common infectious diseases.* International Journal of Epidemiology, 1999. **28**(4): p. 764-8.

14. Pintar, K.D., et al., *A modified case-control study of cryptosporidiosis (using non-Cryptosporidium-infected enteric cases as controls) in a community setting.* Epidemiology and Infection, 2009. **137**(12): p. 1789-99.

15. Gillespie, I.A., et al., *A case-case comparison of Campylobacter coli and Campylobacter jejuni infection: a tool for generating hypotheses.* Emerging Infectious Diseases, 2002. **8**(9): p. 937-42.

16. Voetsch, A.C., et al., *Analysis of the FoodNet case-control study of sporadic Salmonella serotype Enteritidis infections using persons infected with other Salmonella serotypes as the comparison group.* Epidemiology and Infection, 2009. **137**(3): p. 408-16.

17. Centers for Disease Control and Prevention, *Summary of notifiable diseases - United States, 2009.* Morbidity and Mortality Weekly Reports, 2011. **58**(53).

18. *National Electronic Disease Surveillance System (NEDSS): a standards-based approach to connect public health and clinical medicine.* Journal of Public Health Management and Practice, 2001. **7**(6): p. 43-50.

19. Mead, P.S., et al., *Food-related illness and death in the United States.* Emerging Infectious Diseases, 1999. **5**(5): p. 607-625.

20. Kimura, A.C., et al., *Chicken consumption is a newly identified risk factor for sporadic Salmonella enterica serotype enteritidis infections in the United States: A case-control study in FoodNet sites.* Clinical Infectious Diseases, 2004. **38**: p. S244-S252.

21. Domingues, A.R., et al., *Source attribution of human salmonellosis using a meta-analysis of case-control studies of sporadic infections.* Epidemiology and Infection, 2011: p. 1-11.

22. Marcus, R., et al., *Re-assessment of risk factors for sporadic Salmonella serotype enteritidis infections: a case-control study in five FoodNet sites, 2002-2003.* Epidemiology and Infection, 2007. **135**(1): p. 84-92.

23. Clarkson, L.S., et al., *Sporadic Salmonella enterica serotype Javiana infections in Georgia and Tennessee: a hypothesis-generating study.* Epidemiology and Infection, 2010. **138**(3): p. 340-6.

24. Mermin, J., et al., *Reptiles, amphibians, and human Salmonella infection: A population-based, case-control study.* Clinical Infectious Diseases, 2004. **38**: p. S253-S261.

25. Denno, D.M., et al., *Tri-County Comprehensive Assessment of Risk Factors for Sporadic Reportable Bacterial Enteric Infection in Children.* Journal of Infectious Diseases, 2009. **199**(4): p. 467-476.

26. Younus, M., et al., *Case-control study of disease determinants for non-typhoidal Salmonella infections among Michigan children.* BMC research notes, 2010. **3**: p. 105.

27. Centers for Disease Control and Prevention, *2012 Case Definitions: Nationally Notifiable Conditions Infectious and Non-Infectious Case*, 2012: Atlanta, GA.

28. Friedman, C.R., et al., *Risk factors for sporadic Campylobacter infection in the United States: A case-control study in FoodNet sites.* Clinical Infectious Diseases, 2004. **38 Suppl 3**: p. S285-96.

29. Yoder, J.S. and M.J. Beach, *Cryptosporidium surveillance and risk factors in the United States.* Experimental Parasitology, 2010. **124**(1): p. 31-9.

Tables

Table 1. Number and percentage of cases of cryptosporidiosis, giardiasis, and salmonellosis in 16 states reported to the National Electronic Disease Surveillance System (NEDSS)-Base reporting system, 2003-2010.

	Cryptosporidiosis (*n* = 9,266)		**Giardiasis (*n* = 14,806)**		**Salmonellosis (*n* = 64,071)**	
	No.	%	No.	%	No.	%
Alabama	437	4.7	995	6.7	4139	6.5
Arkansas	130	1.4	376	2.5	1838	2.9
Idaho	801	8.6	1222	8.3	1025	1.6
Maine	272	2.9	1088	7.3	610	1.0
Maryland	192	2.1	1361	9.2	4253	6.6
Montana	251	2.7	474	3.2	490	0.8
Nebraska	772	8.3	1049	7.1	1456	2.3
Nevada	93	1.0	257	1.7	438	0.7
New Mexico	656	7.1	554	3.7	1773	2.8
Rhode Island	62	0.7	372	2.5	568	0.9
South Carolina	512	5.5	907	6.1	8129	12.7
Tennessee	457	4.9	1505	10.2	6090	9.5
Texas	3625	39.1	n/a	n/a	25437	39.7
Vermont	353	3.8	1265	8.5	505	0.8
Virginia	512	5.6	3174	21.4	7028	11.0
Wyoming	136	1.5	207	1.4	292	0.5
Total	9,266	100.9*	14,806	99.8	64,071*	100.3*

*Percentages may not equal 100.0% due to rounding

Table 2. Demographics and characteristics of cryptosporidiosis, giardiasis, and salmonellosis cases reported to the National Electronic Disease Surveillance System (NEDSS)-Base reporting system from 16 states, 2003-2010

	Cryptosporidiosis (*n* = 9,266)		**Giardiasis (*n* = 14,806)**		**Salmonellosis (*n* = 64,071)**	
	No.	%	No.	%	No.	%
Age (years)						
Mean (range)	24.6*	(0 – 97)	32.2*	(0 – 95)	26.9	(0 – 107)
Hospitalized	1,362*	18.4	782*	8.1	12,863	29.9
Imported						
Indigenous	4,514*	91.5	3,486*	59.0	27,838	93.4
Out of country	180	3.6	2,068*	35.0	1,044	3.5
Out of jurisdiction	88*	1.8	87*	1.5	257	0.9
Out of state	150*	3.0	266*	4.5	654	2.2
Race						
American Indian	34*	0.6	53	0.7	332	0.8
Asian	62*	1.0	543*	7.3	716	1.7
Black	630*	10.2	1,260*	16.9	5,776	13.9
Native Hawaiian	5	0.1	16*	0.2	31	0.1
White	5,405*	87.8	5,470*	73.5	34,424	83.0
More than 1	14	0.2	17	0.2	70	0.2
Other	3*	0.0	80*	1.1	138	0.3
Hispanic ethnicity	998*	18.7	497*	8.0	8,923	25.3
Sex						
Female	4,651	50.6	6,729*	46.1	32,673	51.6

* p-value < 0.05

Table 3. Frequency, percentage, and adjusted odds ratios of exposure to potential risk factors, cryptosporidiosis and salmonellosis cases reported to the National Electronic Disease Surveillance System (NEDSS)-Base reporting system, 16 states, 2003-2010

	Cryptosporidiosis (*n* = 9,266)		**Salmonellosis (*n* = 64,071)**			
	No.	%*	No.	%*	**aOR****	**(95% CI)**
Animal contact†	3,001	53.7	14,539	51.3	0.99	(0.92, 1.07)
Livestock	456	15.2	595	4.1	2.79	(2.34, 3.32)
Reptile/amphibian	28	0.9	778	5.4	0.13	(0.08, 0.22)
Poultry	127	4.2	669	4.6	0.73	(0.57, 0.95)
Missing/not specified	*3,682*		*28,359*			
Associated with day care	856	13.1	3,020	8.3	1.42	(1.27, 1.58)
Missing/not specified	*2,712*		*27,527*			
Attends day care	658	10.9	2,533	8.2	1.13	(1.00, 1.28)
Missing/not specified	*3,234*		*33,188*			
Lives with day care attendee	383	7.0	774	2.8	2.42	(2.05, 2.86)
Missing/not specified	*3,765*		*36,410*			
Works at day care	82	1.4	119	0.4	3.56	(2.44, 5.17)
Missing/not specified	*3,287*		*33,234*			
Drink untreated water	4,396	92.9	632	2.8	2.57	(2.15, 3.06)
Missing/not specified	*4,534*		*41,105*			
Home water source						
Municipal	4,609	79.7	22,615	82.9	1.00	
Private well	860	14.9	3,035	11.1	1.19	(1.06, 1.33)
Other	212	3.7	812	3.0	0.97	(0.79, 1.19)
None	105	1.8	817	3.0	0.44	(0.34, 0.58)
Missing/not specified	*3,480*		*36,792*			
Home well treatment						
Any	230	53.5	676	59.0	0.96	(0.71, 1.29
None	200	46.5	471	41.1	1.00	
Missing/not specified	*430*		*1,888*			
School/work water source						
Municipal	3,259	88.2	13,332	90.0	1.00	
Private well	108	2.9	207	1.4	1.50	(1.09, 2.08)
Other	177	4.8	530	3.6	1.02	(0.82, 1.27)
None	149	4.0	749	5.1	0.71	(0.57, 0.90)
Missing/not specified	*5,573*		*49,253*			
School/work well treatment						
Any	19	41.3	48	57.1	0.31	(0.08, 1.24)
None	27	58.7	36	42.9		
Missing/not specified	*62*		*123*			
Know other ill persons	2,048	35.4	4,845	16.3	2.73	(2.51, 2.96)
Missing/not specified	*3,487*		*34,366*			
New pet	226	9.7	1,539	13.5	0.71	(0.59, 0.85)
Missing/not specified	*6,947*		*52,693*			
Recreational water exposure†	2,844	49.2	3,353	12.6	6.53	(6.00, 7.12)

Treated	2,089	73.5	1,871	55.8	8.80	(7.96, 9.73)
Untreated	971	34.1	1,489	44.4	4.77	(4.25, 5.37)
Missing/not specified	*3,482*		*26,540*			
Travel	1,473	25.5	4,830	16.9	1.61	(1.47, 1.75)
Missing/not specified	*3,487*		*35,574*			

*Percentages were calculated using cases with nonmissing answers as the denominator

**aOR: Odds ratio adjusted for site, sex, race, and age

†Exposure categories not mutually exclusive; sum of percentages may not equal 100.

Table 4. Frequency, percentage, and adjusted odds ratio of exposure to potential risk factors, giardiasis and salmonellosis cases reported to the National Electronic Disease Surveillance System (NEDSS)-Base reporting system, 16 states, 2003-2010

	Giardiasis (*n* = 23,264)		**Salmonellosis (*n* = 64,071)**			
	No.	%*	No.	%*	**aOR**	**(95% CI)**
Animal contact†	3,159	36.1	14,539	51.3	0.91	(0.84, 0.99)
Livestock	196	6.2	595	4.1	1.03	(0.81, 1.32)
Reptile/amphibian	24	0.76	778	5.4	0.17	(0.10, 0.28)
Poultry	129	4.08	669	4.6	0.69	(0.53, 0.90)
Missing/not specified	*6,059*		*28,359*			
Associated with day care	692	10.3	3,020	8.3	1.34	(0.17, 0.53)
Missing/not specified	*8,083*		*27,527*			
Attends day care	503	6.7	2,533	8.2	1.07	(0.92, 1.25)
Missing/not specified	*7,316*		*33,188*			
Lives with day care attendee	396	4.9	774	2.8	2.06	(1.72, 2.46)
Missing/not specified	*6,753*		*36,410*			
Works at day care	66	0.9	119	0.4	2.73	(1.79, 4.17)
Missing/not specified	*7,257*		*33,234*			
Drink untreated water	788	7.9	632	2.8	5.35	(4.56, 6.28)
Missing/not specified	*4,812*		*41,105*			
Home water source						
Municipal	5,213	63.4	22,615	82.9	1.00	
Private well	1,076	13.1	3,035	11.1	1.12	(1.00, 1.26)
Other	212	2.6	812	3.0	0.93	(0.73, 1.17)
None	89	1.1	817	3.0	0.56	(0.40, 0.78)
Missing/not specified	*6,590*		*36,792*			
Home well treatment						
Any	308	58.0	676	59.0	0.93	(0.68, 1.28)
None	223	42.0	471	41.1	1.00	
Missing/not specified	*545*		*1,888*			
School/work water source						
Municipal	3,157	28.1	13,332	90.0	1.00	
Private well	129	1.1	207	1.4	1.18	(0.84, 1.68)
Other	142	1.3	530	3.6	0.93	(0.70, 1.22)
None	160	1.4	749	5.1	0.75	(0.57, 0.99)
Missing/not specified	*3,588*		*49,253*			
School/work well treatment						
Any	30	49.1	48	57.1	0.26	(0.05, 1.39)
None	31	50.8	36	42.9	1.00	
Missing/not specified	*68*		*123*			
Know other ill persons	1,737	22.7	4,845	16.3	1.36	(1.23, 1.50)
Missing/not specified	*7,140*		*34,366*			
New pet	378	3.1	1,539	13.5	1.13	(0.96, 1.34)
Missing/not specified	*2,598*		*52,693*			
Recreational water exposure†	1,793	20.3	3,353	12.6	2.98	(2.70, 3.28)

Treated	667	37.2	1,871	55.8	2.21	(1.93, 2.52)
Untreated	1,166	65.0	1,489	44.4	3.74	(3.32, 4.22)
Missing/not specified	*5,995*		*26,540*			
Travel	2,946	38.0	4,830	16.9	2.26	(2.01, 2.47)
Missing/not specified	*7,047*		*35,574*			

*Percentages were calculated using cases with nonmissing answers as the denominator

**aOR: Odds ratio adjusted for site, sex, race, and age

†Exposure categories not mutually exclusive; sum of percentages may not equal 100.

Figures

Figure 1. Age distributions for cases of cryptosporidiosis, giardiasis, and salmonellosis in 16 states reported to the National Electronic Disease Surveillance System (NEDSS)-Base reporting system, 2003-2010

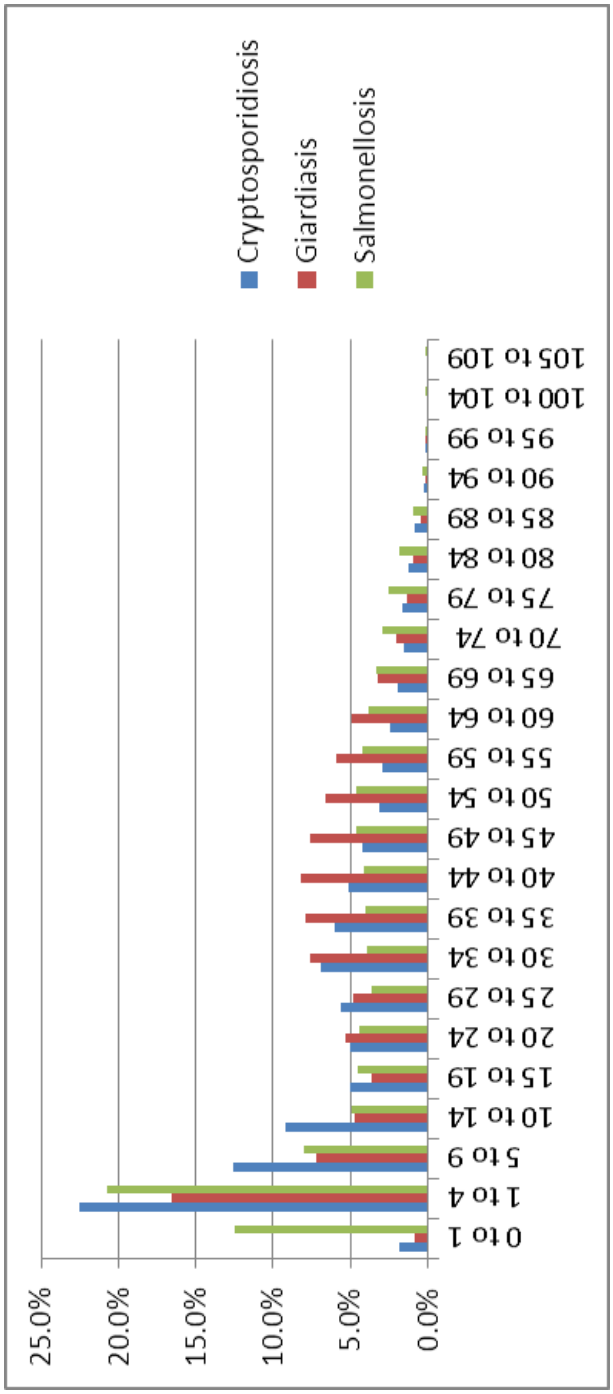
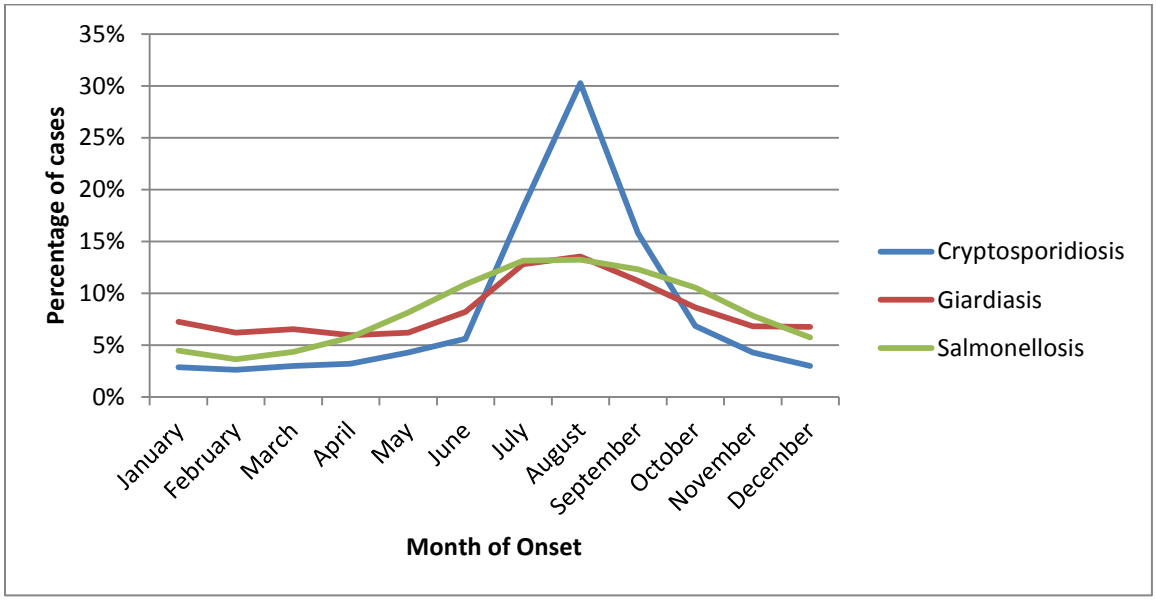


Figure 2. Seasonality of cases of cryptosporidiosis, giardiasis, salmonellosis with known month of illness onset, National Electronic Disease Surveillance System (NEDSS)-Base reporting system, 16 states, 2003-2010



Chapter III: Summary, Public Health Implications, Possible Future Directions

This study used national enteric disease surveillance data and a case-case study design to identify risk factors for cryptosporidiosis and giardiasis infections. A case-case study design could enable timely analyses of suspected outbreaks and expand the use of case previously collected surveillance data. Use of these data might lead to more rapid identification of outbreak cases and timely implementation of control measures. Expanded use of these surveillance data would also provide evidence for the attribution of these diseases to specific exposures, informing science-based prevention efforts.

Cryptosporidiosis and giardiasis are well-known waterborne pathogens, so it follows that the most significant risk factors identified in this study from 2003 to 2010 were water-related. The most significant risk factor for cryptosporidiosis infection was treated recreational water exposure, while giardiasis was closely associated with drinking untreated water. The association between cryptosporidiosis and working in childcare settings has not been previously established and warrants further investigations.

Procedures for following up with cases vary from state to state. Some states may have interviewed all reported cases, while others may only investigate select cases. States also use different questionnaires to assess exposure to risk factors. The high proportion of missing data for some risk factors is likely due to variation in exposure question by pathogen.

As this study demonstrates, a case-case design and surveillance data can be a cost- and resource-efficient method for states to investigate outbreaks. By using available surveillance data as a comparison, states can decrease the time to identifying implicated

sources in outbreak settings. These data can also be used to investigate the proportional contribution of certain exposures (e.g. food, water, animals) to the burden of enteric disease. However, standard guidelines are needed for data collection and management in order to improve the utility of this method.

Appendix

```
*************************************************************************;
**Risk Factors for Cryptosporidiosis and Giardiasis Infections in the ;
**United States: A Case-Case Comparison in 16 States                   ;
**Ellyn Marder                                                         ;
**Master of Public Health, Epidemiology                                ;
**Committee Chair: Patrick Sullivan, PhD, DVM                          ;
**Committee Member: Jonathan Yoder, MSW, MPH                           ;
*************************************************************************;

*********************************************************;
*SC'S CODE FOR DEALING WITH MULTIPLE OBSERVATIONS          ;
*REC WATER, RACE, & ANIMAL TABLES                         ;
*********************************************************;

libname ellyn "\\cdc\project\NCEZID_WDPB_Share\Projects\Yoder,
Jonathan";
*importing the csv files;
/*copying and pasting the code sas generated using the import wizard-
SC*/
PROC IMPORT OUT= WORK.animal
            DATAFILE=
"\\cdc\project\NCEZID_WDPB_Share\Projects\Collier,
Sarah\Ellyn Marder Thesis\animal.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.recwater
            DATAFILE=
"\\cdc\project\NCEZID_WDPB_Share\Projects\Collier,
Sarah\Ellyn Marder Thesis\rec water.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
PROC IMPORT OUT= WORK.race
            DATAFILE=
"\\cdc\project\NCEZID_WDPB_Share\Projects\Collier, Sarah\Ellyn Marder
Thesis\subject race.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
proc sort data=work.race;
      by current_case_data_uid;
run;
proc sort data = work.animal;
      by current_case_data_uid;
run;
proc sort data = work.recwater;
      by current_case_data_uid;
run;
proc sort data = ellyn.fdd_case_data_02292012;
```

```
      by current_case_data_uid;
run;
%macro count(libname, table);
data work.&table.2; set &libname..&table.;
by current_case_data_uid;
if first.current_case_data_uid then do;
count=0;
end;
count+1;
if last.current_case_data_uid then output;
run;
proc freq data=work.&table.2;
title "number of id repeats for &table.";
table count;
run;
%mend count;
%count(work, race);
%count(work, animal);
%count(work, recwater);
%count(ellyn, fdd_case_data_02292012);
%macro contents(libname,table);
proc contents data=&libname..&table.;
run;
%mend contents;
%contents(work, race);
%contents(work, animal);
%contents(work, recwater);
options macrogen; *turning on macrogen so that the log will reflect
what the macro code resolves to;
*recoding rec water table, to produce a dataset with one row per ID and
no ID repeats. Creating 3 new rec water type variables:
untreatedwater treatedwater unknownwatertype;
data work.recwater3;
*making 6 new variables for rec water source (max # of ID repeats is 6
for this dataset);
length RWSource1-RWSource6 $22.;  retain RWSource1-RWSource6;
set work.recwater;
by current_case_data_uid;
*using retain statement and by statement to propagate rec water
information
from the earlier repeats of that id, more info on retain statement
here:
http://www.ats.ucla.edu/stat/sas/modules/collapse2.htm;
*making a count variable to indicate number of repeats for each ID;
if first.current_case_data_uid then do;
*settting new rec water source variables to missing;
RWSource1=' ';
RWSource2=' ';
RWSource3=' ';
RWSource4=' ';
RWSource5=' ';
RWSource6=' ';
*count variable;
count=0;
end;
count+1;
```

```
*populating new rec water source variables with info from each ID
repeat;
%macro recode (iter);
      if count=&iter. then do;
      RWSource&iter.= rec_water_exposure_type_cd_desc_;
  end;
%mend recode;
%recode (1);
%recode (2);
%recode (3);
%recode (4);
%recode (5);
%recode (6);
*using new rec water source variables to make summary rec water type
variables;
if last.current_case_data_uid then do;
RecWaterContact=1;
/*untreatedwater =0;
treatedwater=0;*/
unknownwatertype=0;
array RWSource RWSource1-RWSource6;
do i=1 to 6;
      if RWsource[i] in('hot spring','lake-pond-river-stream','ocean')
then  untreatedwater=1;
      if RWsource[i] in('hot tub-whirlpool-jacu','interactive
fountain','recreational water par','swimming pool',

'vacapool','privpool','privpoo','camppool','compool','edupool','homepoo
l','hosppool', 'kidpool')
                                  then  treatedwater=1;
      if RWsource[i] ='user specified rec wat' then unknownwatertype=1;
end; drop i;
end;
if last.current_case_data_uid then output;*keeping only last ID repeat
for each ID;
run;
data work.animal2; set work.animal;
/*poultry2=0;
livestock2=0;
ReptileAmph2=0;*/
if index(upcase(user_specified_animal_type),'TURTLE')>0
or index(upcase(user_specified_animal_type),'LIZARD')>0
or index(upcase(user_specified_animal_type),'DRAGON')>0 or
index(upcase(user_specified_animal_type),'SNAKE')>0
or index(upcase(user_specified_animal_type),'IGUANA')>0 then do;
ReptileAmph2=1;
end;
if index(upcase(user_specified_animal_type),'CATTLE')>0 or
index(upcase(user_specified_animal_type),'COW')>0
or index(upcase(user_specified_animal_type),'SHEEP')>0 or
index(upcase(user_specified_animal_type),'GOAT')>0
or index(upcase(user_specified_animal_type),'LAMB')>0 then do;
Livestock2=1;
end;
if index(upcase(user_specified_animal_type),'TURKEY')>0 or
index(upcase(user_specified_animal_type),'CHICKEN')>0
```

```
or index(upcase(user_specified_animal_type),'DUCK')>0 or
index(upcase(user_specified_animal_type),'GOOSE')>0 then do;
Poultry2=1;
end;
run;
data work.animal3;
*making 4 new variables for animal source (max # of ID repeats is 4 for
this dataset);
length Animal1-Animal4 $33.;  retain Animal1-Animal4;
set work.animal2;
by current_case_data_uid;
*using retain statement and by statement to propagate animal
information
from the earlier repeats of that id, more info on retain statement
here:
http://www.ats.ucla.edu/stat/sas/modules/collapse2.htm;
*making a count variable to indicate number of repeats for each ID;
if first.current_case_data_uid then do;
*settting new animal variables to missing;
Animal1=' ';
Animal2=' ';
Animal3=' ';
Animal4=' ';
*count variable;
count=0;
end;
count+1;
*populating new rec water source variables with info from each ID
repeat;
%macro recode (iter);
      if count=&iter. then do;
      Animal&iter.= animal_type_cd_desc_txt;
  end;
%mend recode;
%recode (1);
%recode (2);
%recode (3);
%recode (4);
*using new animal variables to make summary animal type variables;
if last.current_case_data_uid then do;
animalcontact=1;
/*poultry1=0;
livestock1=0;
ReptileAmph1=0;
OtherSpecAnimal=0;
poultry=0;
livestock=0;
ReptileAmph=0;*/
array Animal Animal1-Animal4;
do i=1 to 4;
      if Animal[i] in('chicken','turkey') then  poultry1=1;
      if Animal[i] in('cattle','sheep','goats')
                                 then  livestock1=1;
      if Animal[i] in ('lizards','lizard','turtle') then
ReptileAmph1=1;
      if Animal[i] ='user specified animal type' then
OtherSpecAnimal=1;
```

```
end; drop i;
end;
if poultry1=1 or poultry2=1 then poultry=1;
if livestock1=1 or livestock2=1 then livestock=1;
if ReptileAmph1=1 or ReptileAmph2=1 then ReptileAmph=1;
if last.current_case_data_uid then output;*/*keeping only last ID
repeat for each ID;
run;
proc freq data=work.animal3;
table poultry livestock ReptileAmph OtherSpecAnimal animal1-animal4;
run;
proc freq data=work.index;
table cattle2;
run;
data ellyn.recwaterfinal; set work.recwater3;
run;
data ellyn.animalfinal; set work.animal3;
run;
proc sort data=work.race;
by current_case_data_uid;
run;
proc freq data=work.race;
title "number of entries with race=Unknown";
table race_cd_desc_txt;
run;
*counting how many repeats are in the dataset;
data work.racecount; set work.race;
by current_case_data_uid;
where race_cd_desc_txt not in ('Unknown', ' ');
if first.current_case_data_uid then do;
count=0;
end;
count+1;
if last.current_case_data_uid then output;
run;
proc freq data=work.racecount;
title "number of id repeats for racecount";
table count;
run;
data work.race4;
*making 3 new variables for race (max # of ID repeats is 3 for this
dataset);
length Race1-Race3 $75.;  retain Race1-Race3;
set work.race (where=(race_cd_desc_txt not in ('Unknown', ' ')));
*58,376 people were dropped from this dataset - 58,375 had "unknown" as
their only race value and one had
missing as their only race value;
by current_case_data_uid;
*using retain statement and by statement to propagate race information
from the earlier repeats of that id, more info on retain statement
here:
http://www.ats.ucla.edu/stat/sas/modules/collapse2.htm;
*making a count variable to indicate number of repeats for each ID;
if first.current_case_data_uid then do;
*settting new rec water source variables to missing;
Race1=' ';
Race2=' ';
```

```
Race3=' ';
*count variable;
count=0;
end;
count+1;
*populating new rec water source variables with info from each ID
repeat;
%macro recode (iter);
      if count=&iter. then do;
      Race&iter.= race_cd_desc_txt;
  end;
%mend recode;
%recode (1);
%recode (2);
%recode (3);
if last.current_case_data_uid then output;*keeping only last ID repeat
for each ID;
run;
proc freq data=work.race4;
table race1-race3;
run;
data ellyn.racefinal; length RaceFinal $75.; set work.race4;
if race2 ne ' ' then racefinal='Multiple Races';
if race2=' ' then do;
racefinal=race1;
end;
run;
*comparing the final table with race1-race3;
proc freq data=race4;
table race1-race3;
run;
proc freq data=racefinal;
table racefinal/missing;
run;
*end SC code;


*****************************************;
*MERGING ALL TABLES TO MAKE A MASTER DATASET ;
*****************************************;

*creating a libname for the dataset saved on EM Emory H-drive;
libname ellyn '\\dataserver.sph.emory.edu\EMARDE3\My Documents\Thesis -
FINAL\Thesis\SAS';

*sorting the datasets;
proc sort data=ellyn.racefinal;
      by current_case_data_uid;
run;
proc sort data = ellyn.animalfinal;
      by current_case_data_uid;
run;
proc sort data = ellyn.recwaterfinal;
      by current_case_data_uid;
run;
proc sort data = marder.fdd_case_data_02292012;
      by current_case_data_uid;
```

```
run;

*merging the additional tables, corrected for multiple observations;
data ellyn.othertables;
      merge ellyn.racefinal ellyn.animalfinal ellyn.recwaterfinal ;
      by current_case_data_uid;
run;

proc sort data=ellyn.othertables;
      by current_case_data_uid;
run;

data ellyn.alltables;
      merge ellyn.othertables marder.fdd_case_data_02292012 (in=a);
      by current_case_data_uid;
      if a; *this code limits the dataset only to IDs that are in the
fdd_case_data dataset. I'm not sure why there would be IDs that are in
the rec water
      or race table but not in the main dataset -- maybe test entries
or entries that were already cleaned out?;
run;

*checking the code using SC's count coding;
data work.new; set ellyn.alltables;
by current_case_data_uid;
if first.current_case_data_uid then do;
count=0;
end;
count+1;
if last.current_case_data_uid then output;
run;
proc freq data=work.new;
title "number of id repeats for dataset name";
table count;
run;

proc freq data=ellyn.alltables;
      tables condition_cd;
      where nnd_reporting_state_cd_desc_txt = 'Texas' and condition_cd
= '11570';
run;

*************************************;
*CLEANING THE MASTER DATASET              ;
*************************************;
data ellyn.conditions;
      set ellyn.alltables;
      *creating a new dataset with only the observations with
conditions of interest - as cases or controls;
      if condition_cd = '10240' or condition_cd = '10270' or
condition_cd = '10470' or condition_cd = '10530' or condition_cd =
'10640' or condition_cd = '11040'
            or condition_cd = '11540' or condition_cd = '11541' or
condition_cd = '11542' or condition_cd = '11545' or condition_cd =
'11550' or condition_cd = '11565'
```

```
            or condition_cd = '11575' or condition_cd = '11645' or
condition_cd = '12020' or condition_cd = 'D5-41655' or condition_cd =
'DD-80300' or
            condition_cd = 'DD-8480F' or condition_cd = 'DD-95820' or
condition_cd = 'DD-95910' or condition_cd = 'DE-11340' or condition_cd
= 'DE-60046' or
            condition_cd = 'DE-35700' or condition_cd = 'DE-38090' or
condition_cd = '11010' or condition_cd = '11020' or condition_cd =
'11560' or
            condition_cd = '11562' or condition_cd = '11563' or
condition_cd = '11564' then delete;
      *renaming conditions and creating case/control indicator
variables;
      if condition_cd = '11000' then do;
            condition = 'Salmonella';
            Ccase = 0;
            Gcase = 0; end;
      if condition_cd = '11570' then do;
            condition = 'Giardia';
            Gcase = 1; end;
      if condition_cd = '11580' then do;
            condition = 'Crypto';
            Ccase = 1; end;
      *making a new calculated age in years variable;
      if age_reported_unit_cd = 'Y' then calc_age = age_reported;
      if age_reported_unit_cd = 'M' then calc_age = age_reported/12;
      if age_reported_unit_cd = 'W' then calc_age = age_reported/52;
      if age_reported_unit_cd = 'D' then calc_age =
age_reported/365.25;
      if age_reported_unit_cd = 'U' then calc_age = .;
      *Sarah's code for calculating age from birthdate and illness
onset date;
      numDate = datepart(birth_dt);
    format numDate ddmmyy10.;
    strDate = put(datepart(birth_dt),ddmmyy10.);
      birthyr= substr(strdate,7,4);
      birthmth = substr(strdate,4,2);
      birthday = substr(strdate,1,2);
      SASbirth = mdy(birthmth, birthday, birthyr);
      format SASbirth ddmmyy10.;
      numillDate = datepart(illness_onset_dt);
    format numillDate ddmmyy10.;
    illdate = put(datepart(illness_onset_dt),ddmmyy10.);
      illyr= substr(illdate,7,4);
      illmth = substr(illdate,4,2);
      illday = substr(illdate,1,2);
      SASill = mdy(illmth, illday, illyr);
      format SASill ddmmyy10.;
      if SASbirth ne . and SASill ne . then do;
            difference = datdif(SASbirth, SASill, 'act/act'); end;
      calc_age2 = round(difference/365.25, 1);
      *comparing the two calculated age variables;
      if calc_age and calc_age2 ne . then do;
            age_diff = calc_age2 - calc_age;
            end;
      else age_diff = .;
      if age_diff ne . then do;
```

```
            if age_diff le 1 and age_diff ge -1
            then equalage = 'Y';
            else equalage = 'N';
            end;
      *creating a final age variable that is equal to reported age, if
available, and calculated age if not;
      if equalage = 'Y' then age = calc_age;
      if calc_age ne . then do;
            if equalage = 'N' then age = calc_age;
            end;
      if calc_age = . then do;
            if equalage = 'N' then age = calc_age2;
            end;
      label age = 'Age, in years';
      *making 5 year age groups;
      if age ne . then do;
            age04 = 0;  age59 = 0; age1014 = 0; age1519 = 0; age2024 =
0; age2529 = 0; age3034 = 0; age3539 = 0; age4044 = 0; age4549 = 0;
age5054 = 0; age5559 = 0; age6064 = 0; age6569 = 0;
            age7074 = 0; age7579 = 0; age8084 = 0; age8589 = 0; age9094
= 0; age9599 = 0; age100104 = 0; age105109 = 0; end;
      if 0 le age lt 5 then age04 = 1;
      if 5 le age lt 9 then age59 = 1;
      if 10 le age lt 15 then age1014 = 1;
      if 15 le age lt 20 then age1519 = 1;
      if 20 le age lt 25 then age2024 = 1;
      if 25 le age lt 29 then age2529 = 1;
      if 30 le age lt 35 then age3034 = 1;
      if 35 le age lt 40 then age3539 = 1;
      if 40 le age lt 45 then age4044 = 1;
      if 45 le age lt 50 then age4549= 1;
      if 50 le age lt 55 then age5054 = 1;
      if 55 le age lt 60 then age5559 = 1;
      if 60 le age lt 65 then age6064 = 1;
      if 65 le age lt 70 then age6569 = 1;
      if 70 le age lt 75 then age7074 = 1;
      if 75 le age lt 80 then age7579 = 1;
      if 80 le age lt 85 then age8084 = 1;
      if 85 le age lt 90 then age8589 = 1;
      if 90 le age lt 95 then age9094 = 1;
      if 95 le age lt 100 then age9599 = 1;
      if 100 le age lt 105 then age100104 = 1;
      if 105 le age lt 110 then age105109 = 1;
      *splitting out those under 1year for second age group;
      *making 5 year age groups;
      if age ne . then do;
            age01 = 0; age14 = 0;  age59 = 0; age1014 = 0; age1519 = 0;
age2024 = 0; age2529 = 0; age3034 = 0; age3539 = 0; age4044 = 0;
age4549 = 0; age5054 = 0; age5559 = 0; age6064 = 0; age6569 = 0;
            age7074 = 0; age7579 = 0; age8084 = 0; age8589 = 0; age9094
= 0; age9599 = 0; age100104 = 0; age105109 = 0; end;
      if 0 le age lt 1 then age012 = 1;
      if 1 le age lt 5 then age142 = 1;
      if 5 le age lt 9 then age592 = 1;
      if 10 le age lt 15 then age10142 = 1;
      if 15 le age lt 20 then age15192 = 1;
      if 20 le age lt 25 then age20242 = 1;
```

```
	if 25 le age lt 29 then age25292 = 1;
	if 30 le age lt 35 then age30342 = 1;
	if 35 le age lt 40 then age35392 = 1;
	if 40 le age lt 45 then age40442 = 1;
	if 45 le age lt 50 then age45492 = 1;
	if 50 le age lt 55 then age50542 = 1;
	if 55 le age lt 60 then age55592 = 1;
	if 60 le age lt 65 then age60642 = 1;
	if 65 le age lt 70 then age65692 = 1;
	if 70 le age lt 75 then age70742 = 1;
	if 75 le age lt 80 then age75792 = 1;
	if 80 le age lt 85 then age80842 = 1;
	if 85 le age lt 90 then age85892 = 1;
	if 90 le age lt 95 then age90942 = 1;
	if 95 le age lt 100 then age95992 = 1;
	if 100 le age lt 105 then age1001042 = 1;
	if 105 le age lt 110 then age1051092 = 1;
	*making a new variable for month of onset from illness onset
date;
	numMonth = datepart(illness_onset_dt);
	format numMonth ddmmyy10.;
	stronset = put(datepart(illness_onset_dt), ddmmyy10.);
	onsetyr = substr(stronset,7,4);
	onsetmth = substr(stronset,4,2);
	onsetday = substr(stronset,1,2);
	SASonset = mdy(onsetmth, onsetday, onsetyr);
	format SASonset ddmmyy10.;
	*renaming variables;
	state = nnd_reporting_state_cd_desc_txt;
	*dropping OR and DC because not consistent;
	if state = 'Oregon' then delete;
	if state = 'District of Colu' then delete;
	*dropping giardia cases in TX because not reportable;
	if state = 'Texas' and condition = 'Giardia' then delete;
	sex = curr_sex_cd;
	if curr_sex_cd = 'U' then sex = ' ';
	if sex = 'F' then sex_ind = 1;
	if sex = 'M' then sex_ind = 0;
	onsetmonth = onsetmth;
	imported = imported_cd_desc_txt;
	if imported_cd_desc_txt = 'Unknown' then imported = ' ';
	if imported = 'Indigenous' then do;
		indigenous = 1; oocountry = 0; oojuris = 0; oostate = 0;
end;
	if imported = 'Out of country' then do;
		oocountry = 1; indigenous = 0; oojuris = 0; oostate = 0;
end;
	if imported = 'Out of jurisdiction' then do;
		oojuris = 1; oocountry = 0; indigenous = 0; oostate = 0;
end;
	if imported = 'Out of state' then do;
		oostate = 1; oocountry = 0; oojuris = 0; indigenous = 0;
end;
	if hospitalized = 'NI' then hosp = 'N';
	else hosp = hospitalized;
	if hosp = 'Y' then hospital = 1;
	if hosp = 'N' then hospital = 0;
```

```
        race = race_cd_desc_txt;
        if race_cd_desc_txt = 'Other Race' then race = ' ';
        if hispanic_cd_desc_txt = 'N' then hispanic = 'N';
        if hispanic_cd_desc_txt = 'Y' then hispanic = 'Y';
        if hispanic = 'Y' then hispanic_ind = 1;
        if hispanic = 'N' then hispanic_ind = 0;
        if animal_new_pet = 'N' then newpet = 'N';
        if animal_new_pet = 'Y' then newpet = 'Y';
        if newpet = 'Y' then new_pet = 1;
        if newpet = 'N' then new_pet = 0;
        if animal_contact_ind = 'N' then animal_contact = 0;
        if animal_contact_ind = 'Y' then animal_contact = 1;
        if day_care_attend = 'N' then daycareattend = 'N';
        if day_care_attend = 'Y' then daycareattend = 'Y';
        if day_care_work = 'N' then daycareworker = 'N';
        if day_care_work = 'Y' then daycareworker = 'Y';
        if day_care_live = 'N' then daycarelive = 'N';
        if day_care_live = 'Y' then daycarelive = 'Y';
        if associated_with_day_care = 'N' or associated_with_day_care =
'NI' then daycareassoc = 'N';
        if associated_with_day_care = 'Y' then daycareassoc = 'Y';
        if daycareassoc = 'Y' then daycare_association = 1;
        if daycareassoc = 'N' then daycare_association = 0;
        if daycareattend = 'Y' then daycare_attend = 1;
        if daycareattend = 'N' then daycare_attend = 0;
        if daycarelive = 'Y' then daycare_live = 1;
        if daycarelive = 'N' then daycare_live = 0;
        if daycareworker = 'Y' then daycare_worker = 1;
        if daycareworker = 'N' then daycare_worker = 0;
        if home_water_source_cd = 'MUNI' then homewatersource = 0;
        if home_water_source_cd = 'NONE' then homewatersource = 2;
        if home_water_source_cd = 'OTH' then homewatersource = 3;
        if home_water_source_cd = 'PW' then homewatersource = 1;
        if home_well_treat_cd = 'NONE' then homewelltreatment = 0;
        if home_well_treat_cd = 'DINF' then homewelltreatment = 1;
        if home_well_treat_cd = 'FLT' then homewelltreatment = 2;
        if home_well_treat_cd = 'FLTDINF' then homewelltreatment = 3;
        if school_work_water_source_cd = 'MUNI' then otherwatersource =
0;
        if school_work_water_source_cd = 'NONE' then otherwatersource =
2;
        if school_work_water_source_cd = 'OTH' then otherwatersource = 3;
        if school_work_water_source_cd = 'PW' then otherwatersource = 1;
        if school_work_well_treat_cd = 'NONE' then otherwelltreatment =
0;
        if school_work_well_treat_cd = 'DINF' then otherwelltreatment =
1;
        if school_work_well_treat_cd = 'FLT' then otherwelltreatment = 2;
        if school_work_well_treat_cd = 'FLTDINF' then otherwelltreatment
= 3;
        if homewatersource = 1 then home_pw = 1;
        if homewatersource = 0 then home_pw = 0;
        if homewatersource = 2 then home_none = 1;
        if homewatersource = 0 then home_none = 0;
        if homewatersource = 3 then home_oth = 1;
        if homewatersource = 0 then home_oth = 0;
        if otherwatersource = 1 then other_pw = 1;
```

```
      if otherwatersource = 0 then other_pw = 0;
      if otherwatersource = 2 then other_none = 1;
      if otherwatersource = 0 then other_none = 0;
      if otherwatersource = 3 then other_oth = 1;
      if otherwatersource = 0 then other_oth = 0;
      *treatment: none vs. any;
      if homewelltreatment = 0 then home_treat = 0;
      if homewelltreatment = 1 or homewelltreatment = 2 or
homewelltreatment = 3 then home_treat = 1;
      if otherwelltreatment = 0 then other_treat = 0;
      if otherwelltreatment = 1 or otherwelltreatment = 2 or
otherwelltreatment = 3 then other_treat = 1;
      if drink_untreated_water_ind = 'N' then drinkuntreated = 'N';
      if drink_untreated_water_ind = 'Y' then drinkuntreated = 'Y';
      if other_ill_reported_ind = 'N' then knowotherill = 'N';
      if other_ill_reported_ind = 'Y' then knowotherill = 'Y';
      if travel_prior_to_onset_ind = 'N' then travelprior = 'N';
      if travel_prior_to_onset_ind = 'Y' then travelprior = 'Y';
      transmissionmode = transmission_mode_cd_desc_txt;
      if drink_untreated_water_ind = 'N' then drinkuntreated = 'No';
      if drink_untreated_water_ind = 'U' then drinkuntreated = ' ';
      if drink_untreated_water_ind = 'Y' then drinkuntreated = 'Yes';
      if drinkuntreated = 'Y' then drink_untreated = 1;
      if drinkuntreated = 'N' then drink_untreated = 0;
      if rec_water_exposure_ind = 'N' then recwater_exp = 0;
      if rec_water_exposure_ind = 'Y' then recwater_exp = 1;
      if other_ill_reported_ind = 'N' then knowotherill = 'No';
      if other_ill_reported_ind = 'U' then knowotherill = ' ';
      if other_ill_reported_ind = 'Y' then knowotherill = 'Yes';
      if knowotherill = 'Y' then know_ill = 1;
      if knowotherill = 'N' then know_ill = 0;
      if travel_prior_to_onset_ind = 'N' then travelprior = 'No';
      if travel_prior_to_onset_ind = 'U' then travelprior = ' ';
      if travel_prior_to_onset_ind = 'Y' then travelprior = 'Yes';
      if travelprior = 'Y' then travel_prior = 1;
      if travelprior = 'N' then travel_prior = 0;
      if onsetyr = '.' then onsetyr = ' ';
      if animal_contact = . then do;
            livestock = .;
            reptileamph = .;
            poultry = .;
            end;
      if animal_contact = 0 then do;
            livestock = 0;
            reptileamph = 0;
            poultry = 0;
            end;
      if recwater_exp = . then do;
            untreatedwater = .;
            treatedwater = .;
            end;
      if recwater_exp = 0 then do;
            untreatedwater = 0;
            treatedwater = 0;
            end;
      *creating race indicator variables;
      if racefinal ne ' ' then do;
```

```
            race_AI = 0; race_asian = 0; race_black = 0; race_mix = 0;
race_hawaiian = 0; race_other = 0; race_white = 0; end;
      if racefinal = 'American Indian or Alaska' then race_AI = 1;
      if racefinal = 'Asian' then race_asian = 1;
      if racefinal = 'Black or African American' then race_black = 1;
      if racefinal = 'Multiple Races' then race_mix = 1;
      if racefinal = 'Native Hawaiian or Other' then race_hawaiian = 1;
      if racefinal = 'Other Race' then race_other = 1;
      if racefinal = 'White' then race_white = 1;
      *restricting states by reported dates;
      *entering the go live dates for each state;
      if state = 'Alabama' then golive = '1jan07'd;
      if state = 'Arkansas' then golive = '1jun08'd;
      if state = 'Idaho' then golive = '10jan05'd;
      if state = 'Kentucky' then golive = '16jan11'd;
      if state = 'Maryland' then golive = '3apr06'd;
      if state = 'Maine' then golive = '6jul06'd;
      if state = 'Nebraska' then golive = '13jan03'd;
      if state = 'New Mexico' then golive = '3jan06'd;
      if state = 'Nevada' then golive = '3jan05'd;
      if state = 'Rhode Island' then golive = '7dec06'd;
      if state = 'South Carolina' then golive = '1apr04'd;
      if state = 'Tennessee' then golive = '19apr04'd;
      if state = 'Texas' then golive = '10may04'd;
      if state = 'Virginia' then golive = '3jan05'd;
      if state = 'Vermont' then golive = '13dec04'd;
      if state = 'West Virginia' then golive = '23mar11'd;
      if state = 'Wyoming' then golive = '4dec06'd;
      if state = 'Montana' then golive = '23oct06'd;
      format golive ddmmyy10.;
      strreport = put(datepart(last_chg_dt), ddmmyy10.);
      reportyr = substr(strreport,7,4);
      reportmth = substr(strreport,4,2);
      reportday = substr(strreport,1,2);
      SASreport = mdy(reportmth, reportday, reportyr);
      format SASreport ddmmyy10.;
      *deleting cases that were reported before a state's go live date;
      if golive ne . and SASreport ne . then do;
      reportdiff = datdif(golive, SASreport, 'act/act'); end;
      if reportdiff ge 0 then reportvalid = 'yes';
      if reportdiff lt 0 then reportvalid = 'no';
      if reportvalid = 'no' then delete;
      *deleting cases that aren't in the correct timeframe;
      if reportyr = '2011' and onsetyr = '.' then delete;
      if reportyr = '2011' and onsetyr = '2011' then delete;
      if reportyr = '2003' then delete;
      if onsetyr = '1955' or onsetyr = '1987' or onsetyr = '1997' or
onsetyr = '1998' or onsetyr = '2000'
            or onsetyr = '2001' or onsetyr = '2002' then delete;
      *cleaning case status and restricting to 'confirmed' cases;
      if current_case_status_cd ne 'C' then delete;
run;

*************************************;
*OBTAINING FREQUENCIES BY CONDITION     ;
*************************************;
```

```
proc sort data=ellyn.conditions;
      by condition;
run;

proc univariate data=ellyn.conditions;
      var age;
      by condition;
run;

proc freq data=ellyn.conditions;
      tables age04 age59 age1014 age1519 age2024 age2529 age3034
age3539 age4044 age4549 age5054 age5559 age6064 age6569 age7074 age7579
age8084 age8589 age9094 age9599 age100104
               age105109;
      by condition;
run;

proc freq data=ellyn.conditions;
      tables age012 age142 age592 age10142 age15192 age20242 age25292
age30342 age35392 age40442 age45492 age50542 age55592 age60642 age65692
age70742 age75792
               age80842 age85892 age90942 age95992 age1001042
age1051092;
      by condition;
run;

proc freq data=ellyn.conditions;
      tables onsetmonth / list;
      by condition;
run;

proc freq data=ellyn.conditions;
      tables condition;
run;

proc freq data=ellyn.conditions;
      tables hispanic imported state sex hosp racefinal
animal_contact_ind newpet daycareattend daycareworker daycarelive
homewatersource homewelltreatment
      otherwatersource otherwelltreatment drinkuntreated recwaterexp
knowotherill travelprior daycareassoc treatedwater untreatedwater;
      by condition;
run;

proc freq data=ellyn.conditions;
      tables recwater_exp treatedwater untreatedwater;
      by condition;
run;

proc freq data=ellyn.conditions;
      tables animal_contact_ind livestock reptileamph poultry;
      by condition;
run;

proc freq data=ellyn.conditions;
      tables state;
      by condition;
```

```
run;

*************************************************************************;
*STATISTICS: T-TEST/CHI-SQUARE P-VALUES AND ADJUSTED ODDS RATIOS        ;
*************************************************************************;

*switch out Gcase and Ccase in the following code to obtain stats for
both crypto and giardia cases;
proc ttest data=ellyn.conditions;
      class Gcase;
      var age;
run;

*getting chisq p-values for demographic characteristics;
proc freq data=ellyn.conditions;
      tables (age04 age59 age1014 age1519 age2024 age2529 age3034
age3539 age4044 age4549 age5054 age5559 age6064 age6569 age7074 age7579
age8084 age8589 age9094 age9599
                  age100104 age105109 hispanic_ind hospital sex_ind
race_ai race_asian race_black race_mix race_hawaiian race_other
race_white
                  indigenous oocountry oojuris oostate)*Ccase / chisq;
run;

*obtaining aORs;
proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other animal_contact;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other livestock;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other reptileamph;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other poultry;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other daycare_association;
run;

proc logistic data=ellyn.conditions descending;
      class state;
```

```
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other daycare_attend;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other daycare_live;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other daycare_worker;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other drink_untreated;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other home_pw;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other home_oth;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other home_none;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other home_treat;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other other_pw;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other other_oth;
run;
```

```
proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other other_none;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other other_treat;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other know_ill;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other new_pet;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other recwater_exp;
run;

proc logistic data=ellyn.conditions descending;
      class state racefinal;
      model Gcase = state sex_ind age racefinal treatedwater;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other untreatedwater;
run;

proc logistic data=ellyn.conditions descending;
      class state;
      model Gcase = state sex_ind age race_ai race_asian race_black
race_mix race_hawaiian race_other travel_prior;
run;
```