Did anyone get sick this weekend? A pilot study evaluating the effectiveness of using social media posts by recreational watersport users for public health surveillance

By:

Calvin Tan[1], Dale Chen[2], Linda Dix-Cooper[3], Lorraine McIntyre[4]

A Project Submitted in Partial Fulfillment of the Requirement for the Degree of Bachelor of Technology in Environmental Health

British Columbia Institute of Technology

April 2019

1.Lead Author, School of Health Sciences, British Columbia Institute of Technology, 3700 Willingdon Ave, Burnaby, BC V5G3H2

2. Supervisor, School of Health Sciences, British Columbia Institute of Technology, 3700 Willingdon Ave, Burnaby, BC V5G 3H2

3. Collaborator, British Columbia Centre for Disease Control, 655 West 12th Avenue, Vancouver, BC V5Z 4R4

4.Supervisor, British Columbia Centre for Disease Control, 655 West 12th Avenue, Vancouver, BC V5Z 4R4

"The views expressed in this paper are those of the author and do not necessarily reflect the official policy, position or views of BCIT, the Environmental Health Program or its faculty."

ABSTRACT

Background

Recreational water illnesses are not as well known as food borne illnesses in the media. There are several pathogens associated with ingesting surface water including *Giardia*, *Cryptosporidium*, and *Toxoplasmosis*. The use of technology for public health surveillance is also little known to the public and can provide much insight into other illnesses on social media not otherwise reported to public health and medical professionals. Illnesses on social media could represent a portion of unreported cases. These cases could be found on social media as a popular outlet for individual expression.

Methods

Social media posts of illnesses were found using a variety of keywords including symptoms of significant waterborne illnesses and terms associated with human and environmental contamination. Social media posts were collected from forums and popular social media platforms using search terms such as "Sinus", *"Kite Surfing", "Swimming", "Itchy", "Scratchy",* and *"Illness"*. The posts were then correlated with the closest beach water quality data leading up to that date from a nearby location from the incident site. Resulting data around the Columbia river region was 30 posts aggregated into 10 data points.

Results

Social media posts of illnesses and the State of Oregon beach water quality data collected from the Columbia river region were positively correlated. The correlation coefficient of 0.2335 indicates that there is a positive but statistically insignificant correlation between social media posts and beach water quality data. Numerous limitations may have impacted the correlation coefficient. Keywords associated with symptoms were more effective in obtaining threads related to illness and posts compared to other terms.

Conclusions

Manual gathering of social media data for public health surveillance is found to be inefficient and impractical. It remains to be seen whether correlating posts about illness on social media to water quality data is an effective method of surveillance for public health.

Keywords: "Kite Surfing", "Swimming", "Itchy", "Scratchy", "Illness", "Infection", "Vomiting", "Sick", "Sewage", "Sinus", "Bacteria", "Social media"

Table of Contents

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Foodborne and waterborne illnesses continue to be one of the most prevalent and easily preventable public health issues in Canada (The Chief Public Health Officer's Report on the State of Public Health in Canada 2013 – Food-borne and water-borne infections, 2013); Underreporting of these diseases hinders the allocation of resources, underestimates the severity of the issue, and hinders disease transmission prevention programs. Despite modern advancements in treating surface water, a weekend of watersports can still quickly turn from “wicked sick” to “bathroom sick” due to the presence of pathogens in outdoor recreational waterways.

Among possibilities that contaminate recreational water include fecal contamination from wild animals, boaters, and chemical contamination from untreated effluent (Edwards & Crozier, Introduction: Water Supplies and Health, 2017). These factors may be contributing to illness in recreational water users. In the past decade, social media has become popular for sharing information, leading one to wonder if it can be used for public health surveillance. The null hypothesis for this project is that there is no correlation between the social media posts and the water quality data. The alternative hypothesis is that there is a correlation between the social media posts and the water quality data.

Project statement

This study was conducted to investigate whether online social media and websites review can provide insight into unreported illnesses of an environmental and public health nature. This study will also endeavor to determine if environmental contamination can be associated with an increase in reportable illnesses by correlating beach sample reports with forum posts or blog entries on social media platforms. This project will also investigate if certain keywords or websites where information regarding waterborne or general ailments of public health concern can be used to gather data. An emphasis will be placed on outdoor recreational water users who experienced symptoms and reported them on social media platforms.

BACKGROUND

Waterborne illnesses background

Diseases associated with drinking water

Waterborne illnesses are described as "Enteric infections spread through fecal contamination of drinking water (Edwards & Crozier, Introduction: Water Supplies and Health, 2017). The organisms of most concern with waterborne illnesses continues to be *Giardia* and *Cryptosporidium*; these two organisms cause the diseases Giardiasis and Cryptosporidiosis, respectively. Both organisms are protozoans that produce cysts and oocysts as part of their life cycle (Edwards & Crozier, Introduction: Water Supplies and Health, 2017). Cysts and oocysts are dormant forms of the protozoa which become vegetative once environmental conditions allow for their growth. Because cysts and oocysts cannot be destroyed by chlorination of water, modern water treatment methods include a filtration step to remove the cysts and oocysts (Edwards & Crozier, Water Treatment Processes, 2018).

Toxoplasmosis is another disease caused by protozoa that can humans can get infected by from exposure to water. Although outbreaks are rare, toxoplasmosis can be in contaminated waters and can cause severe nervous and respiratory infections as well as death (Chiu, 2018). Infants and pregnant are especially vulnerable (Chiu, 2018). Those infected with toxoplasmosis are often asymptomatic or otherwise experience flu like symptoms with "swollen lymph glands or muscle aches and pains that last for a month or more" (Parasites - Toxoplasmosis (Toxoplasma infection), 2018). Sequelae of toxoplasmosis tends to refer to newborns and can include neurological impairment or death (Chiu, 2018).

Giardiasis tends to last one to two weeks with symptoms such as diarrhea, gas, floating greasy stools, abdominal cramps, nausea, and dehydration (Parasites - Giardia, 2015). Long term consequences, or sequelae, of Giardiasis can include ocular complications, arthritis, skin allergies or myopathy (Halliez, 2013). Cryptosporidiosis also tends to last one to two weeks with symptoms such as watery diarrhea, stomach cramps, dehydration, nausea, vomiting, fever, and weight loss (Parasites - Cryptosporidium, 2015). Debilitating symptoms such as these can have a negative impact on human health and causes economic, social, and mental unrest among other issues. Sequelae of Cryptosporidiosis include weight loss, abdominal pain, diarrhea, eye pain, joint pain, and fatigue (Stiff E Rhianwen, 2017).

The danger of these protozoa is that it only takes 10 cysts of *Giardia* to cause Giardiasis in humans (Pathogen Safety Data Sheets: Infectious Substances – Giardia lamblia, 2011). Further, it only takes as low as 1-5 oocysts of *Cryptosporidium* to cause illness (Pathogen Safety Data Sheets: Infectious Substances – Cryptosporidium parvum, 2011). Toxoplasmosis can be caused by 10 oocysts (Pathogen Safety Data Sheets: Infectious Substances – Toxoplasma gondii, 2014). In 2016, the BCCDC reported an incidence rate of 11.4 cases of Giardiasis, 2.4 cases of Cryptosporidiosis, and 0.0 cases of Toxoplasmosis per 100,000 people in BC (Reportable Disease Dashboard, n.d.). Considering the low infective dose of these protozoa, the incidence rate, and the ease at which they can be consumed, it is clear there is a threat to public health and wellbeing.

Notable outbreaks associated with waterborne diseases

Waterborne illnesses can cause disease to the point that it exceeds the average illness incident rate in the population. In Canada, *Giardia*, *Cryptosporidium*, and *Toxoplasma* have all been known to cause outbreaks in the past.

An outbreak of giardiasis occurred in the Niagara public health district in Canada in January 2013 (Jordan pool reopened after Giardia outbreak, 2013). Seventy-five people fell ill after contracting Giardia at a pool (Jordan pool reopened after Giardia outbreak, 2013). The outbreak was identified to have started as early as October 2012, but the implicated location was not identified until January 2013 (Jordan pool reopened after Giardia outbreak, 2013).

Waterborne cryptosporidiosis was implicated in an outbreak that occurred in Saskatchewan in 2001 (Chiu, 2018). It was estimated that 5800-7100 people were affected by this outbreak with 275 lab confirmations of Cryptosporidium as the causative agent (Chiu, 2018). Surface water contamination was most likely to have been the source of the outbreak as there was a "malfunction of solid unit contact and increased turbidity" (Chiu, 2018).

An outbreak of toxoplasmosis occurred in the Greater Victoria region of British Columbia in March 1995 (R, et al., 1997). One hundred individuals were identified to have been affected by the outbreak (R, et al., 1997). Although no "conventional source" was implicated, it was suspected that one reservoir had caused the outbreak after a period of high rainfall and turbid water (R, et al., 1997). The researchers concluded that a municipal system that uses unfiltered chloraminated surface water caused the outbreak (R, et al., 1997).

Recreational water usage as a risk factor for waterborne illness

Recreational water refers to rivers, lakes and coastal waters and can be used for activities such as swimming, surfing, water skiing, white water sports, underwater diving, sailing, boating and shellfish gathering (About recreational water quality and health, 2017). Intentional or incidental immersion of the body, including the head, into surface waters increases the risk of becoming ill (Macleod, 2018).Swallowing recreational surface water also increases the chances of becoming ill. Recreational water bathers have an increased risk of becoming ill, having symptoms of gastrointestinal illness, and experiencing ear ailments compared to non-bathers (Leonard, Singer, Ukoumunne, Gaze, & Garside, 2018).

When illnesses occur in recreational water bathers, because the symptoms overlap with other types of enteric illnesses, the true cause of the symptoms may be obscured. Not everyone goes to a doctor if they become ill. Lab analysis of clinical and environmental samples is the best way to confirm symptoms, to identify a causative organism, and to attribute the source to recreational water over other routes of transmission, such as person to person contact, food, etc. Lab sampling is seldom ordered by doctors though and although lab testing is the best way to determine the cause of illness, there is a small chance that false negatives results are reported instead. Epidemiological surveillance is a tool that can help capture and identify those that do fall ill but did not get reported. The true number of illnesses is often underestimated even with surveillance; however, it gives a better overall understanding of how much enteric illness actually happens and how often people are affected.

Sources of recreational water contamination

Sources of contamination related to human activities

Contamination of water can occur from point sources or non-point sources (Edwards & Crozier, Introduction: Water Supplies and Health, 2017). Point sources refer to “pollution coming from a single identifiable source” while non-point sources refer to “pollution resulting from multiple sources” (Edwards & Crozier, Introduction: Water Supplies and Health, 2017). Pollution and turbidity issues with water can come in the form of wastewater discharges, runoff from watersheds, algal decomposition products, and high iron content (Edwards & Crozier, Surface Drinking Water Sources, 2017). Additional sources include “Malfunctioning private sewage disposal systems, storm water

outfalls, bottom sediments, and agricultural drainage (Macleod, 2018). To ensure the safety and quality of water for recreational use, routine sampling is often done by health authorities, sewage plant personnel, or contractors.

Guidelines and regulations for wastewater discharges

Currently, the guidelines set out by Health Canada for recreational water state that there must be no more 400 *E. coli*/100mL in a single sample or no more than 200 *E. coli*/100mL in an average of five samples (Guidelines for Canadian Recreational Water Quality, 2012). This testing is done to look for indicator organisms, like *E. coli*, that can be in the water when there is gross contamination which can cause illness. Indicator organisms are "easily detectable organisms whose presence correlates directly to one or more pathogens contaminating an environment" (Ikner, 2018). The presence of indicator organisms often means that there are also pathogens in the water that can cause illness.

Animal and environmental contamination

Environmental factors relating to contamination of waters includes "geographical factors, pollution volume, characteristics of the waste, rainfall, prevailing winds, and thermal stratification" (Macleod, 2018). These factors generally disturb the waters and potentially mixing any contaminants in that may have already been present in the waters. Cyanobacteria can also be a potential contaminant. "Cyanobacteria are bacteria that share features of both bacteria and algae" (Macleod, 2018). Forty-six species of cyanobacteria, also commonly known as blue green algae, produce toxins which humans may ingest, inhale, or contact by skin (Macleod, 2018).

Animal contamination commonly occurs when manure or fecal matter gets into recreation waters via rainwater runoff, direct fecal inputs, or leaching into groundwater (About recreational water quality and health, 2017). Manure can contain fecal matter from livestock such as "dairy cattle, beef cattle, deer and sheep" (About recreational water quality and health, 2017). It is often difficult to tell if water pollution was caused by animals because it is a non-point source contamination (Fewtrell & Kay, 2015). Bird and dog visitations can often account for an increase in microbial load of up to 10^5 cfu/day for birds and 10^{10} cfu/day for dogs (Fewtrell & Kay, 2015). This would account for natural bird visitations and pets that water users might bring to beaches.

However, most notable is that through the retrospective study of previous papers, the researchers came to the conclusion that "there

was no evidence for associations between swimming-associated gastrointestinal illness and exposure to natural recreational water polluted with feces from non-human sources" (Fewtrell & Kay, 2015). This has high implications because, even accounting for numerous other factors, agricultural activities seem to have no association in recreational water users and an increase in waterborne related illnesses.

Utilizing social media for public health data collection

Background about public health in social media

Recently, public health professionals have looked at using social media as a method of monitoring cases and outbreaks of disease. Platforms such as Facebook and Twitter are examples of platforms where health professionals can gather their data (Vance, Howe, & P.Dellavalle, 2009). Other prospective platforms include Snapchat, Instagram, reddit, and WeChat. Social media data collection is not limited to just these sites though. Technology and innovation are making great advances and new social media platforms are popping up all the time. Usage of social media for public health data collection does not appear to be openly discussed. Research into incorporating social media for surveillance and outbreak management suggest great potential for the practice; however, limitations include lack of studies into the use of social media for data collection and methods of practical use even if there are significant findings (Charles-Smith, et al., 2015).

In September 2008, the CDC collaborated with Google in an attempt to "compare volumes of flu-related search activity against reported incidence rates for the illness displayed graphically on a map" (Schmidt, 2012). This collaboration created Google Flu Trends and it started as a result of findings that "spikes in flu queries and disease outbreaks often coincide" (Schmidt, 2012). This project allowed the CDC to monitor for potential flu outbreak to allow for early intervention. In cases where social media does closely predict "influenza-like illnesses", some of the terms that were useful include "home worse," "cough night," "sore head," "influenza," "symptom," "shortage," "hospital," and "infection" (Schmidt, 2012). These terms reportedly predicted flu outbreaks up to two weeks ahead of CDC surveillance data (Schmidt, 2012). Some enteric diseases also have symptoms such as fever, nausea, and vomiting, making it difficult to differentiate between the two. Thus, it may be possible to use similar keywords to look for posts of waterborne illnesses.

Existing methods of online public health surveillance

One existing software that can mine for data is HealthMap. The software "mines government websites, social networks and local news reports to map potential disease outbreaks" (Brown, 2015). The software was able to map out the Ebola outbreak using this data nine days before the World Health Organization declared the epidemic (Brown, 2015). Usage of HealthMap and historical data that may be kept could prove to be useful in finding any correlational evidence between social media and outbreaks. Furthermore, the existence of software like HealthMap could serve as a model for user friendly, easy to use future software that could provide useful data to health officials.

The public health department in Chicago has teamed up with a group named Smart Chicago to develop an application that analyzes tweets about food poisoning (Brown, 2015). The city increased inspections and enforcement on offending establishments as a result of the findings (Brown, 2015). New York City's department of health and mental hygiene also teamed up with Columbia University and Yelp to find reviews about food poisoning (Brown, 2015).

Other methods of using technology would include data scraping and machine learning to find data in an efficient manner. "Data scraping refers to a technique in which a computer program extracts data from output generated from another program. Data scraping is commonly manifest in web scraping, the process of using an application to extract valuable information from a website" (What is Data Scraping?, n.d.). "Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention" (Machine Learning, n.d.). The combination of these two concepts allows professionals in computer science to create software for health officials. The process of looking through individual websites and posts and interpreting the data for positive results could be minimized greatly and streamlined to utmost efficiency.

One existing application of machine learning used for public health surveillance is found in a machine-learned model created to monitor and detect foodborne illnesses in real time (Sadilek Adam, 2018).

The model, FINDER, is 3.1 times more likely to find a restaurant unsafe during inspections compared to existing methods and is more reliable in identifying and implementing corrective actions that would otherwise lead to

a foodborne illness outbreak (Sadilek Adam, 2018).

Limitations associated with social media surveillance

Use of social media for surveillance purposes was also associated with a high potential for both false positives and false negative results (Social Media in Public Health, 2015). False positives for flu outbreaks can occur as seen with the Google Flu Trends (Schmidt, 2012). Online search behavior may not necessarily be an indicator for flu outbreaks (Schmidt, 2012). This occurs due to false positives or false negatives. The reliability of information, the potential inability to identify authors, the large volume of information and the potential for information inaccuracies pose challenges (Social Media in Public Health, 2015).

Benefits of social media surveillance and data collection

Studies find that the uses of social media can also include information dissemination, engaging the public or stakeholders by using social media as a platform, crowdsourcing for data collection, and broadcasting environmental health issues, as well as surveillance and outbreak monitoring (Hempel, 2014). The benefits of using social media to investigate public health issues include timely and accessible information that can be shared globally as long as the information is accurate and reliable (McNab, 2009). It is often suggested that social media data can be used to complement official data that is released later (Rumi Chunara, 2012).

Social media surveillance was correlated with existing surveillance programs in a study and was particularly good at rapid detection of disease trends (Social Media in Public Health, 2015). Data can often be real time and early intervention methods can be more effective than existing methods (Charles-Smith, et al., 2015). “Monitoring data extracted from social media provide insight into the opinions that are at a certain moment salient among the public, which enables public health institutes to respond immediately and appropriately to those public concerns” (Eysenbach, 2015).

Social media is highly accessible due to the prevalence of cellular devices, the internet, and the information that can be accessed (Rumi Chunara, 2012). Although outbreaks are rare, when using social media posts to track them, results may be as accurate as official surveillance data that is released much later by government officials. A study comparing blog and Facebook posts with official reports found that the information matched up well (Schmidt, 2012). This particular study suggests developing “an automatic coding system that captures

content and user's characteristics that are most relevant to the diseases within the National Immunization Program and related public health events" (Eysenbach, 2015).

Despite all the research and benefits of using social media for public health purposes, there appears to be few standardized or publicly available method by which health officials can consistently and frequently use to conduct monitoring or surveillance on social media. Analytical methods appear to be complex and are restricted to the knowledge of few individuals or groups. The information is not easily made commercial and is not user friendly. Most other studies appear to focus on dissemination of information via social media (Rosemary Thackeray, 2012).

Literature Review Summary

Pathogens such as *Giardia* and *Cryptosporidum* are predominant waterborne protozoa which can cause diseases in recreational water users. The symptoms of the diseases that are caused by the protozoa can be used to look for posts and entries about illness. Although microorganisms are naturally present, certain sources of contamination like wastewater treatment plants can cause higher amounts of pathogens to be present in waters. Most literature does not address the methods and means for collection and analysis of data from social media. Data can be done via web scraping and machine learning for large samples but manual collection and analysis is also possible. The prospective use of social media for data gathering can allow for rapid data collection with fairly high accuracy and high accessibility for health officials and the public alike. However, data is subject to higher false positives and false negatives. Reliability, inaccurate information, and large volumes of data to be analyzed can also pose challenges.

METHODOLOGY

This study evaluated the utility of social media posts about water illness among outdoor recreational water users as a potential public health surveillance indicator. Can social media posts about illness be used for public health surveillance effectively? The null hypothesis is that there is no correlation between social media posts about illnesses and beach water quality data. Therefore, the alternate hypothesis is that there is a correlation between social media posts about illnesses and beach water quality data. Inclusion criteria included watersport activities that involve head immersion below water or that put them at risk of swallowing water including swimmers, surfers, kiteboarders, wind surfers, and freediving scuba divers. All post data was

collected retrospectively from the following social media platforms: reddit, Twitter, and forums such as iWindsurf and BayAreaKiteboarding which had a beach location indicated. It was gathered and correlated with the respective location's water sampling data collected from environmental and health officials. Real time data collection is not feasible for data collection at the time of this project. Health authority data was collected primarily to determine if the water was indicative of pathogens that would have caused the illness in recreational water users. Potential contamination sources and general water contamination were gathered from peer reviewed journal papers and government organization websites. Private industry websites were browsed for information regarding wastewater treatment. Computing software to collect and analyze social media posts was explored and found to be beyond the scope of this project.

Materials and Methods

All data was collected from the internet and public social media platforms using a computer with an internet connection. Some social media platforms were previously provided by the project sponsors and were also used. Using these websites, a preliminary search for posts about illness was done using some of the search keywords used throughout the study. By using publicly available social media platforms, this research project avoided the ethical issue of gathering data from private forums. However, by electing to gather data from only public forums, there may have been a missed opportunity for valid and valuable data about illnesses to be gathered from private forums. Methods of gathering and analyzing social media data for the purpose of conducting correlation analysis is scarce or not for public viewing. Expert help from the computer science industry was limited and because of time limitations, an executable program to collect and analyze social media data was not available at the time of this study. As such, social media data and platforms were gathered manually by browsing through results from Google and online watersport communities via internal search engines using search keywords such as "Kite Surfing", "Swimming", "Itchy", "Scratchy", "Illness", "Infection", and "Vomiting" (Dix-Cooper & McIntyre, 2018). Careful data screening to control data quality regarding posts from search terms such as "Sick" was required. The literature review also revealed social media data is often unreliable with false positives, false negatives, and inaccuracies as potential challenges (Social Media in Public Health, 2015). Thus, posts containing these keywords were analyzed a little more carefully.

Screening and inclusion criteria for social media posts

Due to variation in people's preferences in post length, some posts may contain more than one case of illness. In the event that two or more illnesses are reported in one post, the illnesses were counted as single posting but reported as multiple illnesses. The rationale for this is to keep consistency with previously conducted studies that counted number of queries which do not necessarily represent just one person who is ill (Schmidt, 2012). By counting number of illnesses separately, the magnitude of illness was examined and the potential underreporting of illnesses due to aggregation of illnesses was quantified. It is also possible to count illnesses that are not related to recreation water use in any way. The symptoms that were included in such a post were analyzed carefully according to symptoms that would be associated with recreation water illnesses. If the symptoms closely match but it is uncertain that the illness was of waterborne in nature, the post was not be included in the data. To further keep consistency with previous studies, posts about waterborne ailments such as ear and sinus infections commonly associated with water sports were also included in the data. Any applicable posts regarding sequelae from waterborne illnesses were included as well.

Geographical considerations and criteria

We correlated number of social media posts by region to water quality data from health authorities such as Vancouver Coastal, Fraser Health, State of Oregon Water Quality, California Water Boards, and CEDEN. Recreational water sites were selected from a map of common beach locations where windsurfers may launch. We included Canadian, American, and Mexican sites to ensure ample data given the niche sports that were involved.

If health authority data for water quality data were unavailable, an independent provider of microbial data would be used instead, or the microbial data was taken from a location that is most likely for watersport users to launch from while being as representative of the whole geographical area described as possible. This was done because some watersport users do not always enter the waters during the peak summer season from April to September. Most health authorities conduct their water sampling in the same warmer season time range; thus, there is a colder time period from October to March where there may be a lack of water quality data for some geographical sites. To determine the best water sample data to use and to determine the location of the activity, the posts, website, and public profiles were scrutinized for hints. Popular websites such as theswimguide.org listed common activities that

visitors engage in which helped further focus the most likely location where a water user was infected. Some more detailed information may be obtained from those that are more familiar with water sporting in the local area. Google Maps was used to determine the closest sampling site if an exact location could not be determined.

Surfing appears to be associated with "increased incidence of several categories of symptoms, and associations were stronger if surfing took place shortly after rainstorms; higher levels of fecal indicator bacteria were strongly associated with fever, sinus pain/infection, wound infection, and gastrointestinal symptoms within 3 days of rainstorms" (Arnold, et al., Acute Illness Among Surfers After Exposure to Seawater in Dry- and Wet-Weather Conditions, 2017). Therefore, if a post mentioned weather being a factor to the illness, historical weather records for the region of interest were investigated around the time of illness. These particular symptoms were also used to guide inclusion or exclusion from the dataset. If an exact location could not be determined but the general location of the incident was known, the microbial reading that was geographically closest to the incident was used.

The timing of social media posts was also something of consideration in the data collection. Unless it is explicitly specified in a post, it is not known when someone went for their activities and when that person began experiencing symptoms. Also, unless a person made a post immediately after experiencing the first symptom, it is not possible to know for sure if the date of the posting was when they first started experiencing symptoms. This affects the computation of a disease incubation period and also affects determining if an illness was caused by recreational waters and the associated activity or if it was due to other causes. As most posts were screened for recreational water activity and gathered, as much as possible, from online recreational water communities, all screened reported illnesses were assumed to be due to recreational water use. Further, since recall bias must be accounted for in determining illness date and unless a post explicitly states otherwise, it was assumed that the date of the incident will have occurred within the last two weeks of the posting and during rainfall conditions when applicable.

In the analysis of correlating number of postings and beach water quality data from health authorities, it is possible for the data to be biased. If there are two different locations geographically close to each other but there is

only one beach water quality report, it will be necessary to correlate both locations to the same report. Where reports show that a microbiological count is “less than” a certain value, the analysis will be conducted using the highest displayed value of that interval. For example, if a count is reported as “less than 5 MPN/100mL”, the data will use the value of 5 MPN/100mL for the analysis because the true value is not known other than that it was no more than the limit of detection. If a website does not provide whole numbers for water sample data but has sufficient historical evidence to indicate a microbial problem, historical data was analyzed. To remain on the side of caution, an assumption that the site passes the water quality standard will be made. If there is no sampling site at the location of interest, a site as geographically close as possible was chosen with respect to the waterbody and location of incident.

Analytical considerations and criteria

With further regards to location, previous studies collected data in an aggregate manner to allow data to be more comparable between regions (Google Flu Trends Estimates, 2015). To keep data consistent with previous studies and to have data collected for initial analysis, this project also began with aggregate data collection in the initial collection phase (Chen, 2018).

Many symptoms of waterborne illness such as fever and vomiting can also be associated with a number of other illnesses such as the flu or foodborne illnesses. Similarly, by manually screening of posts for waterborne symptoms, it is possible that some posts would be ambiguous in reporting of symptoms or not all symptoms are reported in the post. Thus, it is possible for a true waterborne illness to be excluded. Mislabeling of posts positive for waterborne illness can also be a problem in this study. We may be finding a correlation between water quality data and number of social media postings when there may be no correlation in reality.

Aggregation of data reduced the amount of analysis required and remained consistent with previous studies; however, this may underestimate the true number of illnesses. Since this project is heavily reliant on posts about illnesses, if persons do not report their illnesses online or have no access to the internet, there will be less data as a result. Analysis will be biased towards locations that have sampling data from health authorities. If a post did not have matching health authority sampling data for the location, it was aggregated with another area or excluded which overestimated or underestimated the incidence in each case respectively.

Data was analyzed with statistical software in Microsoft Excel and NCSS 12. For each different recreational water site, social media data was correlated with water quality data over time. The format of the data is numerical; therefore, the data was analyzed by correlation and presented with graphs and tabulations of data. A standard one-tailed alpha value of 0.05 was used to reject the null hypothesis. A beta value of 0.20 was used to determine the probability of rejecting the null hypothesis incorrectly. The beta value was computed through the power value from analysis in NCSS 12. The magnitude of the coefficient of determination was used to determine how well the data could be used to predict analyses of the data in a linear model for that area. Finally, the magnitude of the correlation coefficient was used to determine how strong the linear relationship created by the dataset is (Heacock, 2018). Cut off ranges ranged from 0.00 to 0.25; 0.25 to 0.50; 0.50 to 0.75; and 0.75+ in the positive and negative scale (Heacock, 2018). These cut off ranges described the relationship with varying degrees of descriptors starting from “no relationship” to “very good to excellent relationship” (Heacock, 2018). For the Columbia river dataset, it may be best not to remove any outliers due to the small amount of data points. Outliers were determined by examining the residual plots versus the microbial readings. A test removal of one outlier seen in figure 1 shown below at the coordinates (93,1) may be a potential candidate for removal. This results in a stronger correlation coefficient of 0.4868 but this does not change the conclusion of the result that there was no statistically significant positive correlation.

STATISTICS, DATA, AND RESULTS

Columbia River dataset

The Columbia River is a river from the northern part of the Washington state that runs along the Washington-Oregon border and ends at the Pacific Ocean. This area surrounding the river is also commonly called “The Gorge”. The horizontal stretch of the river that is on the Washington-Oregon border east of Portland was where most of the incidents occurred. See Appendix A and B for maps of the region. Orange markers indicate the general area of an incident location and blue markers indicate the sampling sites as shown in the State of Oregon’s Water Quality Monitoring Data maps (AWQMS - Water Quality Monitoring Data, n.d.).

Table 1. Summary statistics for Columbia River social media posts and water quality data

	E.coli data (MPN/100mL)	Social Media Post data (posts)
Number of data points	10	30
Min	2	1
Max	162	6
Average	38.07	3
Standard deviation	51.22644825	1.56347192

Summary statistics shown in table 1 were analyzed in Microsoft Excel 2016. This table shows some summary statistics of the data values which may be of some interest. The largest *E.coli* microbial value, 162 MPN/100mL. This reading is a potential indicator as high indicator microbial readings are commonly associated with illness. The standard deviation was 51.2 for microbial readings, indicating high variability and values.

Table 2. Summary of keyword usage in finding threads for Columbia River dataset

Keyword	**Threads found**
Infection	1
Sick	1
Diarrhea	3
Sewage	1
Sinus	3
Bacteria	1

Table 2 shows the keywords used in each search instance to find the thread with illness postings. Some threads were found using more than one keyword and the keywords were counted more than once for the purpose of this table. “Diarrhea” and “Sinus” appear to be the keyword most frequently associated postings in the Columbia River region.

Table 3. NCSS 12 Summary statistics for Columbia River correlation of social media posts and water quality data

Run Summary Section

Parameter	Value	Parameter	Value
Dependent Variable	Number_of_posts_about_illness	Rows Processed	10
Independent Variable	Microbiological_value	Rows Used in Estimation	10
Frequency Variable	None	Rows with X Missing	0
Weight Variable	None	Rows with Freq Missing	0
Intercept	2.7287	Rows Prediction Only	0
Slope	0.0071	Sum of Frequencies	10
R-Squared	0.0545	Sum of Weights	10.0000
Correlation	0.2335	Coefficient of Variation	0.5375
Mean Square Error	2.600084	Square Root of MSE	1.612478

Statistics derived from NCSS statistical software version 12.0.9 has 4 significant parameters to explore as shown in table 3. The intercept is 2.72, indicating that if there were no microbial readings, the number of posts would be 2.72. This could be reasonable given that background posts about illness may be due to a number of factors including those unrelated to watersports. The slope of 0.0071 indicates that for the dataset, the number of posts increase by 0.0071 for every 1 increase in MPN/100mL that is made. An R-Squared value of 0.05 seems to indicate that the data poorly fits a straight line and a correlation coefficient shows a weak correlation. A correlation coefficient of 0.2335 indicates a weak positive correlation which may be due to the low number of data points.

Figure 1. Correlation of social media posts from various sources and water quality data from governing bodies for Columbia River in Oregon

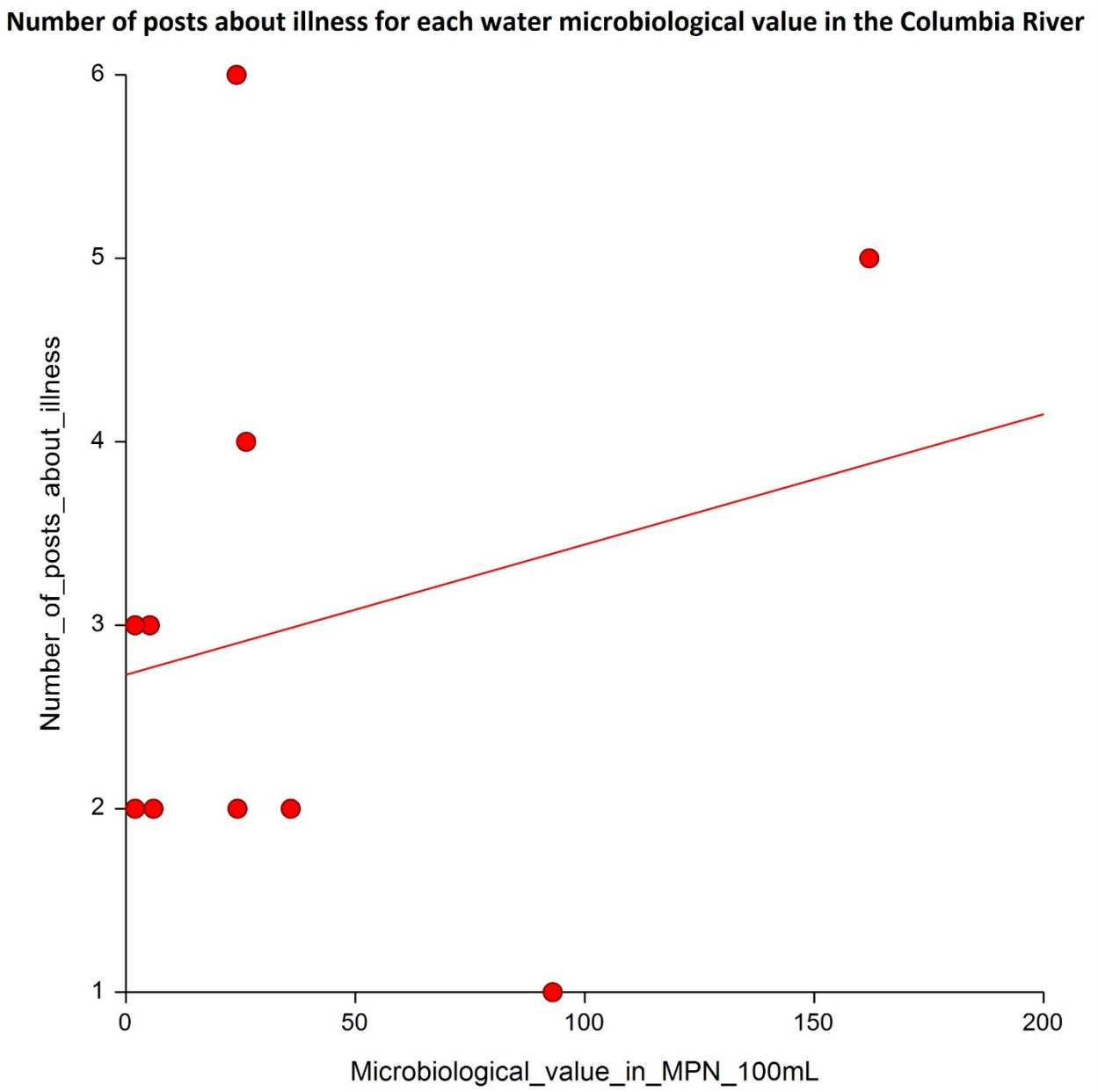

Figure 1 shows the all posts about illness in the Columbia River region with its respective aggregated microbiological value for the region of the incident site.

DISCUSSION

Social media posts were gathered manually from various online communities and groups including reddit, Twitter, and online forums such as iWindsurf and BayAreaKiteboarding. An emphasis was placed on posts containing keywords relating to illness. A correlation analysis was conducted on the posts with corresponding water quality data around the date of an incident.

Columbia River

The Columbia River dataset contained 30 posts categorized into dates respective of water quality data approximated by the social media post. There were 10 data points for correlation. The correlation coefficient for the Columbia River as determined by NCSS 12 was 0.2335. This indicated that there is a weak positive correlation between social media posts about illness and water quality data for the Columbia River region. The laboratory limit of detection was not an issue with determining a fixed whole number for analytical purposes but would be something to consider in future endeavors.

Sampling sites were spread throughout the Columbia river region such that a water quality reading was close by for each incident location except for the point near Rooster Rock; thus, geographical location of the sampling site is less of a confounding factor.

Previous research findings

Using Google to find research papers, there appeared to be no previous studies that that specifically correlate illness in recreational water users to the water quality data of the location they surfed at. Therefore, this study will be difficult to extrapolate and compare. Some studies utilize machine learning to gather and analyze public health surveillance data; however, the scope of the study is focused on foodborne illnesses (Sadilek Adam, 2018). One study does analyze waterborne ailments but the variable factor in that experiment is the weather. Although the weather can indirectly have an impact on the microbiology of the waters by disturbing the sediments, the study only covered surfers and excludes all other recreational water users (Arnold, et al., Acute Illness Among Surfers After Exposure to Seawater in Dry- and Wet-Weather Conditions, 2017).

In these two studies, the previous research found that for the foodborne illness study, the FINDER model used in the study was 3.1 times more likely to deem restaurants unsafe (Sadilek

Adam, 2018). This study also gathered data in an aggregate form. The major difference is that foodborne data was gathered and anonymity of users was preserved (Sadilek Adam, 2018). This study did not use any keyword terms to screen and sort data. The study was conducted in Las Vegas, USA and deemed restaurants unsafe based on inspection reports (Sadilek Adam, 2018). This methodology is similar to this social media study because it is also focused on one geographical location. The study shows the potential that this experiment could have if machine learning was incorporated into the project design. Regression analysis was used to analyze the data in Sadilek's experiment whereas correlation was used in this experiment (Sadilek Adam, 2018).

In the second study that was similar to this experiment, the researchers analyzed fecal contaminants in water after rainstorms to determine if it is the cause of acute illness (Arnold, et al., Acute Illness Among Surfers After Exposure to Seawater in Dry- and Wet-Weather Conditions, 2017). That study used a defined surfer population from San Diego, CA and recorded symptoms of those that were exposed to recreational waters (Arnold, et al., Acute Illness Among Surfers After Exposure to Seawater in Dry- and Wet-Weather Conditions, 2017). This study did not use social media but used a survey to gather this information (Arnold, et al., Acute Illness Among Surfers After Exposure to Seawater in Dry- and Wet-Weather Conditions, 2017). The study found that seawater exposure lead to an "increased incidence rate ratio for all [symptoms or illnesses defined in the experiment] except for upper respiratory illness" (Arnold, et al., Acute Illness Among Surfers After Exposure to Seawater in Dry- and Wet-Weather Conditions, 2017). The symptoms reported by the surfers are similar to those that were found in this social media experiment potentially indicating some degree of agreement for indicators of illness keywords and terms.

Public health significance

The lack of studies into using social media or technology for public health surveillance made this experiment difficult to conduct. This also indicates, however, a gap in knowledge and much potential for growth in what researchers call "machine learned epidemiology" (Sadilek Adam, 2018). Refining currently existing models or initiating development in this unexplored field may yield possible surveillance methods or provide a basis for a future public reporting method for illnesses that is easily accessible.

Search keywords and terms for finding social media data for public health surveillance can be derived from symptoms of illnesses but researchers and officials need to be cognizant

of jargon or slang that online communities may use in place of formal or official terms. This means officials may need more education or communication with the community in order to become familiar with these terms. Meeting and familiarizing with communities that participate in activities that may have potential effects on their health can also help build relationships and understanding for future educational communications.

LIMITATIONS

There are a significant number of limitations to this project. First, research ethics approval was not obtained for this project. This proved to be an issue throughout the project. One notable limitation without research ethics approval was that only public data could be gathered and private groups could not be analyzed. A second limitation was the difficulty in determining the location of the incident and where it was absent, the location of the poster to approximate the incident location. Even given the resources of a map containing common surfing locations; using other posts to determine the location; and public profiles with location information where applicable, determining the exact location of an incident was rather imprecise.

The platform where posts were obtained can also be a limiting factor. The size and activity of the community can affect the representative sample of the true proportion of illnesses that occur. This affects the sample size of the posts gathered and potentially the significance or conclusion of the analysis.

Another problem was the lack of data for social media posts and water quality data alike. There was a notable lack of one or both of Canadian social media data and water quality data. As such, data from the United States was the focus of this project. More social media posts would have made the analysis stronger; however, the bigger problem was with the lack of water quality data. The lack of microbial data was, in fact, a limitation in itself as some incident locations had no microbial data at all nearby. This resulted in a removal of post data or the use of a surrogate water sampling site for analysis.

Some posts had incident dates that could not be matched with an exact data for existing sources of water quality data. This was somewhat corrected by using another water sampling site or a different date for the same sampling site regarding all water quality data. Expanding the geographical area for post searching helped with finding more posts. Less stringent criteria overall helped find more posts as well.

There was some jargon used but using Google to determine the meaning of the terms was adequate. When locations were shortened as jargon, Google proved to be helpful as well as the map of common surfing locations. Overall lack of time and social media platforms with posts were limitations. An expansion of keywords was required partway through the project. There may also have been some bias involved in the data collection since it is more likely for those that are sick to report their condition online. The implication of this bias may result in a disproportionally high incidence rate compared to platforms that may include posts about illness and posts explicitly stating a person is not ill. This is an unavoidable bias, however, due to the nature of the project objectives.

Some of these limitations may have been avoided if research ethics approval was obtained. As this project has to be done manually without the aid of computer programs, more time to complete the project would also have been helpful.

KNOWLEDGE TRANSLATION

This project demonstrates the difficulty of gather data for public health surveillance without the use of more advanced computing methods. As technology develops and is more commercially utilized, the public health sector could easily benefit from greater surveillance and disease reporting by incorporating technological data gathering methods.

In terms of water sampling, there is not much Canadian or American data available for certain incidents or areas for recent years. This may indicate that more water sampling is needed for the purpose of public health surveillance and data gathering. This may be done by trained community enthusiasts or health officials.

FUTURE RESEARCH

The author of this paper, with respect to suggestions by the project sponsors, have the following recommendations for future research projects:

1. Expand on this project with the inclusion of research ethics approval and website scrapping or machine learning
2. Compare regions with higher microbial counts to regions with lower microbial counts and conduct a similar project to compare the regions
3. Conduct a survey, online or in person, to gather data regarding illnesses and correlate with existing water quality data

4. Conduct a similar experiment with food establishment infractions and social media posts about foodborne illness
5. Adjust for differences in population size of water users between sites using incidence rates to estimate the number of illnesses

CONCLUSIONS

Social media posts that were suspected to be about waterborne illness were gathered and compared to water quality data in the Columbia River region in Oregon, USA. Statistical analysis was done by correlation and resulted in a weak correlational value of 0.2335. Manual gathering of social media data for public health surveillance is found to be inefficient and impractical. Further study is required in order to determine the effectiveness of using social media for public health data gathering. It remains to be seen whether correlating posts about illness on social media to water quality data is an effective method of surveillance for public health.

ACKNOWLEDGEMENTS

The lead author of this paper would like to thank the co supervisors, Helen Heacock and Dale Chen, for their advice, feedback, and support throughout this course. Their comments provided guidance in producing results for this project.

STATEMENT OF WORK

Lorraine McIntyre, a Food Safety Specialist with the BC Centre for Disease Control, provided the research project opportunities through collaboration with the BCCDC and the supervisors. Her support in providing project resources in the form of websites, journal articles, and time gave the project direction and scope.

Much gratitude must also be given to Linda Dix-Cooper, a scientist with the BC Children's and Women's hospital. She was essential in providing many resources related to the study population. The feedback she provided helped fine-tuned the writing in this paper.

COMPETING INTEREST

The authors declare that they have no competing interests.

REFERENCES

About recreational water quality and health. (2017, November 18). Retrieved from Environmental Health Indicators New Zealand: http://www.ehinz.ac.nz/indicators/recreational-water/about-recreational-water-quality-and-health/

Arnold, B. F., Schiff, K. C., Ercumen, A., Benjamin-Chung, J., Steele, J. A., Griffith, J. F., . . . Colford Jr., J. M. (2017). Acute Illness Among Surfers After Exposure to Seawater in Dry- and Wet-Weather Conditions. *American Journal of Epidemiology*, 866-875.

Arnold, B. F., Schiff, K. C., Ercumen, A., Benjamin-Chung, J., Steele, J. A., Griffith, J. F., . . . Colford Jr., J. M. (2017). Acute Illness Among Surfers After Exposure to Seawater in Dry- and Wet-Weather Conditions. *American Journal of Epidemiology*, 866-875.

AWQMS - Water Quality Monitoring Data. (n.d.). Retrieved from State of Oregon Department of Environmental Quality: https://orwater.deq.state.or.us/Login.aspx

Brown, J. (2015, January 8). *Using Social Media Data to Identify Outbreaks and Control Disease*. Retrieved from Emergency Management: http://www.govtech.com/em/health/Social-Media-Data-Identify-Outbreaks.html

CANADA-WIDE STRATEGY FOR THE MANAGEMENT. (2014). Retrieved from Canadian Council of Ministers of the Environment: https://www.ccme.ca/files/Resources/municipal_wastewater_efflent/PN_1522_MWWE_Five_Year_Rvw_2014.pdf

Cercarial Dermatitis. (2017, December 28). Retrieved from CDC: https://www.cdc.gov/dpdx/cercarialdermatitis/index.html

Charles-Smith, L. E., Reynolds, T. L., Cameron, M. A., Conway, M., Lau, E. H., Olsen, J. M., . . . Corley, C. D. (2015, October 5). *Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review*. Retrieved from PLOS One: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0139701#abstract0

Chen, D. (2018, November 16).

Chiu, J. (2018). Unit 4 Part 13 Waterborne & Foodborne Protozoans.

Phone correspondance with Dix-Cooper, L., & McIntyre, L. (2018, October 3). (C. Tan, Interviewer)

Edwards, J., & Crozier, V. (2017). Introduction: Water Supplies and Health.

Edwards, J., & Crozier, V. (2017). Surface Drinking Water Sources.

Edwards, J., & Crozier, V. (2018). Water Treatment Processes.

Eysenbach, G. (2015, May 26). *Disease Detection or Public Opinion Reflection? Content Analysis of Tweets, Other Social Media, and Online Newspapers During the Measles Outbreak in the Netherlands in 2013*. Retrieved from NCBI: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4468573/

Fewtrell, L., & Kay, D. (2015, January 22). Recreational Water and Infection: A Review of Recent Findings. *Current*

Environmental Health Reports, pp. 85-94.

Google Flu Trends Estimates. (2015, August 19). Retrieved from Google: https://www.google.com/publicdata/explore?ds=z3bsqef7ki44ac_

Guidelines for Canadian Recreational Water Quality. (2012, August 22). Retrieved from Health Canada: https://www.canada.ca/content/dam/canada/health-canada/migration/healthy-canadians/publications/healthy-living-vie-saine/water-recreational-recreative-eau/alt/pdf/water-recreational-recreative-eau-eng.pdf

Halliez, C. M. (2013). Extra-intestinal and long term consequences of Giardia duodenalis infections. *World Journal of Gastroenterology*, 8974–8985.

Harmful Algal Bloom (HAB)-Associated Illness. (2017, December 13). Retrieved from CDC: https://www.cdc.gov/habs/illness.html

Harmful Algal Bloom (HAB)-Associated Illness. (2017, December 13). Retrieved from CDC: https://www.cdc.gov/habs/illness-symptoms-marine.html

Heacock, H. (2018, November 1). Module 4C - ANOVA and Correlation Regression.

Hempel, M. (2014, March 9). *The use of social media in environmental health research and communication: an evidence*. Retrieved from NCCEH: http://www.ncceh.ca/sites/default/files/Guelph-Hempel-2014.pdf

How Wastewater is Treated . (n.d.). Retrieved from metrovancouver: http://www.metrovancouver.org/services/liquid-waste/treatment/treatment-plants/how-wastewater-treated/Pages/default.aspx

Ikner, L. (2018). *Water Quality Analysis via Indicator Organisms*. Retrieved from Jove: https://www.jove.com/science-education/10025/water-quality-analysis-via-indicator-organisms

Isaac Chun-Hai Fung, Z. T.-W. (2015). The use of social media in public health surveillance. *Western Pacific Surveillance and Response Journal*, 3-6. Retrieved from World Health Organization Western Pacific Region: https://ojs.wpro.who.int/ojs/index.php/wpsar/article/view/319/500

Jordan pool reopened after Giardia outbreak. (2013, March 6). Retrieved from NiagaraThisWeek.com: https://www.niagarathisweek.com/news-story/3270860-jordan-pool-reopened-after-giardia-outbreak/

Leonard, A. F., Singer, A., Ukoumunne, O. C., Gaze, W. H., & Garside, R. (2018). Is it safe to go back into the water? A systematic review and meta-analysis of the risk of acquiring infections from recreational exposure to seawater. *International Journal of Epidemiology*, 572-586.

Machine Learning. (n.d.). Retrieved from SAS: https://www.sas.com/en_ca/insights/analytics/machine-learning.html

Macleod, M. (2018). Lecture 8 Recreational Waters.

McNab, C. (2009, August). *World Health Organization.* Retrieved from What social media offers to health professionals and citizens: http://www.who.int/bulletin/volumes/87/8/09-066712/en/

MT. HOOD AND THE COLUMBIA RIVER GORGE . (n.d.). Retrieved from https://www.hood-gorge.com/

MUNICIPAL WASTEWATER REGULATION. (2012, April 20). Retrieved from BC Laws: http://www.bclaws.ca/EPLibraries/bclaws_new/document/ID/freeside/87_2012#section96

Parasites - Cryptosporidium. (2015, February 20). Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/parasites/crypto/illness.html

Parasites - Giardia. (2015, July 21). Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/parasites/giardia/illness.html

Parasites - Toxoplasmosis (Toxoplasma infection). (2018, October 2). Retrieved from CDC: https://www.cdc.gov/parasites/toxoplasmosis/gen_info/faqs.html

Pathogen Safety Data Sheets: Infectious Substances – Cryptosporidium parvum. (2011). Retrieved from Government of Canada: https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/pathogen-safety-data-sheets-risk-assessment/cryptosporidium-parvum-pathogen-safety-data-sheet.html

Pathogen Safety Data Sheets: Infectious Substances – Giardia lamblia. (2011). Retrieved from Government of Canada: https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/pathogen-safety-data-sheets-risk-assessment/giardia-lamblia.html

Pathogen Safety Data Sheets: Infectious Substances – Toxoplasma gondii. (2014, September 19). Retrieved from Government of Canada: https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/pathogen-safety-data-sheets-risk-assessment/toxoplasma-gondii-pathogen-safety-data-sheet.html

R, B. W., S, K. A., H, W. D., L, I.-R. J., Alison, B., B, E. S., & A, M. S. (1997). Outbreak of toxoplasmosis associated with municipal drinking water. *The Lancet*, 173-177.

Reportable Disease Dashboard. (n.d.). Retrieved from BC Centre for Disease Control: http://www.bccdc.ca/health-info/disease-system-statistics/reportable-disease-dashboard

Rosemary Thackeray, B. L. (2012, March 26). *Adoption and use of social media among public health departments*. Retrieved from BMC Public Health: https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-12-242?report=reader#Sec3

Rumi Chunara, J. R. (2012). Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak. *The American Journal of Tropical Medicine and Hygiene* , 39-45.

Sadilek Adam, C. S. (2018, November 6). *Machine-learned epidemiology: real-time detection of foodborne illness at scale*. Retrieved from Digital Medicine: https://www.nature.com/articles/s41746-018-0045-1

Schmidt, C. W. (2012, January). *Trending Now: Using Social Media to Predict and Track Disease Outbreaks*. Retrieved from

NCBI: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3261963/

Social Media in Public Health. (2015, January). Retrieved from NCCHPP: http://www.ncchpp.ca/docs/2015_TC_KT_SocialMediaPH_en.pdf

Stiff E Rhianwen, D. P. (2017). Long-term health effects after resolution of acute Cryptosporidium parvum infection: a 1-year follow-up of outbreak-associated cases. *Journal of Medical Microbiology*, 1607-1611.

The Chief Public Health Officer's Report on the State of Public Health in Canada 2013 – Food-borne and water-borne infections. (2013, October 13). Retrieved from Government of Canada: https://www.canada.ca/en/public-health/corporate/publications/chief-public-health-officer-reports-state-public-health-canada/chief-public-health-officer-report-on-state-public-health-canada-2013-infectious-disease-never-ending-threat/food-borne-and-water-

Vance, K., Howe, W., & P.Dellavalle, R. (2009, April). Social Internet Sites as a Source of Public Health Information. *Dermatologic Epidemiology and Public Health*, pp. 133-136. Retrieved from ScienceDirect : https://www.sciencedirect.com/science/article/pii/S0733863508001083?via%3Dihub

What is Data Scraping? (n.d.). Retrieved from CLOUDFLARE: https://www.cloudflare.com/learning/security/threats/data-scraping/

REFERENCE TO COMPLETE PAPER

Tan, C. (2019). Did anyone get sick this weekend?: Did anyone get sick this weekend? A pilot study evaluating the effectiveness of using social media posts by recreational watersport users for public health surveillance. BCIT Environmental Health Journal.