

2016

Real Time Data Analysis of Clostridium defficile and Echerichia coli

Alexa Reyes

Follow this and additional works at: <http://scholarworks.lib.csusb.edu/osr>

 Part of the [Medicine and Health Sciences Commons](#)

Recommended Citation

Reyes, Alexa (2016) "Real Time Data Analysis of Clostridium defficile and Echerichia coli," *OSR Journal of Student Research*: Vol. 4 , Article 2.

Available at: <http://scholarworks.lib.csusb.edu/osr/vol4/iss1/2>

This Article is brought to you for free and open access by CSUSB ScholarWorks. It has been accepted for inclusion in OSR Journal of Student Research by an authorized editor of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.

Real Time Data Analysis of Clostridium defficile and Echerichia coli

ALEXA REYES, Class of 2016, Health Science- Environmental
Mentor: **Monideepa**, Assistant Professor, Health Science & Human Ecology

ABSTRACT:

Background: Infectious diseases have not seized to exist for it has been found that microorganisms like Clostridium defficile and Escherichia coli bacteria are capable of mutating. As a result, antibiotic resistant novel strains, such Clostridium defficile (C.diff) that cause devastating new disease continue to pose a problem. An early detection tool for such emergent strains and the evaluation of the most highly impacted geographic locations is critical to mitigate the burden of such condition. Much of the literature, however, has shown that while archived data has little predictive nature as well as limited geographic analysis capabilities. On the other hand, novel analytic tools, such as Google Trends, can provide insight into a potential outbreak and provide geographic disparities of illnesses relying on search volume. The purpose of this paper was to use Google Trends to evaluate its potential analytical ability for two infectious disease causing agent: C.diff and E.coli. **Methods:** Keywords C. diff and E.coli were used to evaluate search volume using Google Trend analysis. Regions of interest, Unites States and Europe, were systematically evaluated. **Results:** Results showed that in the United States eastern and southern geographic locations, such as Pennsylvania, Maryland, and Massachusetts had a higher search volume for both keywords, when compared to the entire country. Interestingly, these states also report the highest rate of E.coli and C.diff infected patients per year. Countries like Canada, and the United Kingdom also demonstrated a high search volume when compared to China or Russia, further reflecting current trends in such infection. **Conclusion:** Similar to the literature noting the predictive power of Google Trends related to Flu and Lyme disease, the results of this study, demonstrate a potential relationship between search volume for such bacterial diseases and associated incidences across several countries.

KEYWORDS: Google Trends, C.diff, E.coli, Search Volume, Mutations, Emergent Strains, Potential Outbreaks

Historically, infectious diseases have been the leading cause of death affecting the majority of the global population. Pandemics like the Plague of Justinian and the bubonic plague are scarcely an example of the potency bacterium's can have on the human body. Little was known about these deadly microorganisms at that time, and therefore, ineffective medications that would eradicate such diseases were lacking. Microorganisms have claimed the lives of millions over the years bringing humanity to the brinks of significantly low populations that is until the development of antibiotics. We like to believe that civilization has won the fight over most infections leading us to focus our time, energy and resources on chronic diseases not realizing that microorganism are capable of

mutating (1). In fact, industrialized nations, such as United States, New Zealand, Canada etc., are going through the final stages of the epidemiologic transition: the era of antibiotic resistant novel strains that cause devastating diseases previously unknown to man. For example, new strains of bacterium's like Escherichia coli (E.coli) is a clear evidence that mutations are correlated with the development of new antibiotic resistant strains, as is the new emerging Clostridium defficile(C.diff) (2).

Recent empirical evidence further shows that patients with the new strains of such bacteria no longer respond to existing medications, and as a result, thousands of individuals per year have died from these infections worldwide (3). In fact, in 2011 over half a million individuals were

infected with the C.diff bacteria and 29 thousand of them died within 30 days of initial diagnosis. It is possible that these infection causing bacterium's are somewhat a representation of a potential next big outbreak. In recent years however, novel real time data evaluation tools have enabled researchers and public health professionals to track the rising burden of new and emerging health issues, thus providing an early detection tool. As a means of obtaining relative levels of health related trend data over time, Google Trends has increasingly been utilized over the years. For example, a study using Goggle Trend tools demonstrated the ability of such a tool to highlight Lyme Disease trends (4). Interestingly, these results suggested that Northeastern, Northwestern and Midwestern regions of the United States not only had high search traffic but also has a higher susceptibility rate for Lyme disease in comparison to other parts of the country. Thus making it a highly endemic region.

As such, similar strategies could be utilized to evaluate when key geographic areas of the global population and United States are impacted the most with such emerging infectious diseases by searching for real time data on search volume. Moreover, as a result of initiatives funded by the Patient Protection and Affordable Care Act of 2010 (5), health literacy is a major public health issue. Evaluating the search term of the general population can further reveal the patterns of health literacy in the global and United States' population; further providing healthcare professionals the tools to target the most at-risk groups. In this study, we utilized Google Trends to evaluate both C. diff and E.coli search volume, both globally and in the United States to evaluate real time trends.

METHODS

Search trend data was collected for analysis through Google Trends. Key words used were Cdiff, C.diff, Ecoli, and E.coli. Other searches included infections, symptoms and treatment for each of the bacteria. Google trends showed how often these particular terms were entered relative to their search volume across various regions of the globe. It is important to note that the display of data is limited to how Google Trends allows for data visualization, as it

has been done with the previous literature (6). This study was emphasized primarily on data searches in the United States and European countries independently. After narrowing down these regions to specific areas like cities comparisons between countries were eventually made. All Google trend researches with key words C.diff and E.coli were made August 22nd, 2015. On September 7th, 2015 without a period between the letters the following keywords searched were Cdiff and Ecoli.

RESULTS

After the launch of google trends in 2004, there has been a significant increase of C.diff and E.coli google search trends over the years. Focusing on C.diff, regional searches showed that over all the U.S., United Kingdom, and Canada had the only and highest volume of C.diff searches globally. Of these three countries the United States has volume score of 100, while, the United Kingdom has 80 and Canada 50 as observed in Figure 2.

Figure 1. Cdiff Globally over Time

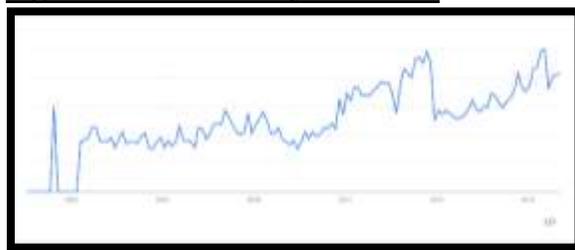


Figure 1 illustrates Google search Cdiff globally over time. The peaks represent all Cdiff searches at a definite time.

Figure 2. C.diff Regional Interest



Figure 2 depicts Google search terms on a global scale utilizing C.diff as the keyword. The columns represent the total amount of C.diff searches in the U.S., United Kingdom, and Canada.

It was found that in this category the rest of the countries shown on google trends like Russia or China had 0 search volumes. However, further investigation showed that starting 2013 Canada and England have had 1a sudden increase in search volume. Google trends news headlines informed that in 2014 Europe Roche launched a program to detect the presence of C.diff in markets that accepts products from la Marca. Such actions could have been taken due to hyper virulence of C.diff strains worldwide which called for improved diagnostics and management strategies. The volume of searches could have potentially increased as a result of recurrence of the disease in infected patients. Narrowing it down to the U.S., it was observed that C.diff search volumes peaked to an all-time high in 2012 compared to 2006 and has slowly surpassed the number of searches in 2015 as shown in Figure 3.

Figure 3. Cdiff US Interest over Time

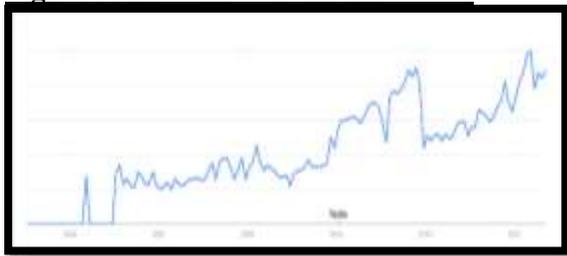


Figure 3 illustrates Google search terms in the US over time. The peaks represent all Cdiff searches at a definite time.

Figure 4. C.diff US Regional Interest



Figure 4 illustrates Google search terms in the U.S. utilizing C.diff as the keyword. The columns represent the total amount of C.diff searches in the country.

Interestingly, when concentrating solely on the U.S. (Figure 4), it was found that the majority of high search volumes are made in the

eastern and southern part of the country. States such as Ohio, Pennsylvania Maryland and Massachusetts appear to have 90 plus volume score for searches and continue to rise. Other states are not falling behind with 80 plus search volumes. As seen in the Figure 3 and 4 provided 2015 has had several peaks on what is still midyear.

Figure 5. Ecoli Global Interest over Time

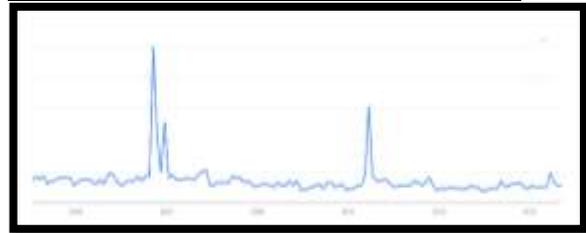


Figure 5 illustrates Google search term globally over time. The peaks represent all Ecoli searches at a definite time.

Figure 6. E.coli Global Regional Interest

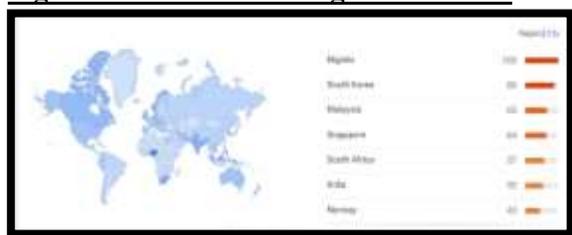


Figure 6 depicts Google search on a global scale utilizing E.coli as the keyword. The columns represent the total amount of E.coli searches.

Upon analyzing google trend results, it was evident that the highest amount of E.coli searches were made in 2006 and 2007 in the U.S as it had the highest peak in Figure 7. There are various known E.coli strains circulating the United States of which O157:H7 is known to be the most frequently acquired bacteria in comparison to other strains (7). Nevertheless, it did not take long for other countries to fall victim to this strain and new emergent strains. In Figure 5, Google trend shows that in late 2011 there was a sudden spike in search volume, area of focus Europe. News headline showed that there was an E.coli outbreak in Germany of an unknown strain. This new emergent and drug resistant E.coli strain is now known to be O104:H4 (7). The following year there was an E.coli outbreak of the strain O157:H7 in both in the U.S. of which affected its agriculture (8). Data shows that like

C.diff most of the outbreaks that befall in American soil typically occur in the Eastern and southern parts of the country.

Figure 7. Ecoli US Interest over Time

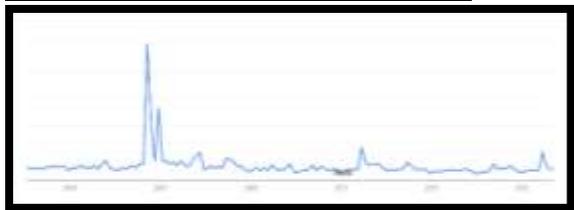


Figure 7 illustrates Google search terms in the U.S. over time. The peaks represent all Ecoli searches at a definite time.

Figure 8. E.coli US Regional Interest



Figure 8 illustrates Google search terms in the U.S. utilizing E.coli as the keyword. The columns represent the total amount of E.coli searched in the country

DISCUSSION

The major findings from this study were a potential relationship between increase search volume and increase incidence of infection. In addition, it was determined that new and/or recurrent cases of C.diff and E.coli infections that mainly occur in southern and eastern regions of the United States tend to relate to the search volume. Alarmingly enough not only have these American bacterium's spread to European countries but developed a mutation; yielding new strains. Similar to the case of Lyme Disease study (4) our results also demonstrated Google Trends as a real time data analytic tool to evaluate potential geographic area and time of an outbreak based on search volumes.

For example, the results of this study showed a sudden peak of C.diff and E.coli search volume in the United States and Europe taking place in 2011 and early 2015. Coincidentally, at the same point in time it was reported that there was an outbreak of these bacteria on these general

areas. Center for Disease Control and Prevention (CDC) reported shiga toxin-producing *E. Coli* O104:h4 to be the primary reason for the European outbreak in 2011 (7). Late 2011 United States CDC opened an inquiry on what was then known to be a multistate outbreak on the E.coli serotype O157:H7 bacterial infections associated to romaine lettuce. Prevalent cases occurred in Arizona, Arkansas, Illinois, Indiana, Kansas, Kentucky, Minnesota, Missouri and Nebraska(9).

Similarly, that same year states like West Virginia, Ohio, Pennsylvania and Tennessee were affected by a C.diff bacterial outbreak. In the year following online C.diff search volume increased exponentially and have remained at an all-time high. Recent studies show have further highlighted increased rates of such emergent strains with higher virulence, resistance, and/or both (10). European countries are no longer the exception. It is known that the Netherlands is now dealing with an increase prevalence of C.diff infections due to new Polymerase Chain reaction Ribotype 078. News of this new outbreak may have been a contributing factor to the increase search volume reported in 2011 and subsequent years. Other contributing factors could be the rise of recurrent cases, increase incidences, frequent outbreaks and online access in these areas.

Recurrent C.diff and E.coli cases have been on the rise, causing illness and in severe cases death. In 2000, the Journal of Clinical Microbiology published a study that focused on recurrent cases of E.coli infections. It was hypothesized that the bacteria can cause repeated respiratory tract infection in patients with severe cystic bronchiectasis. Interestingly enough, there are 71 different E.coli strains that cause recurrent urinary tract infections in women. However, Clostridium defficile relapses can largely be due to endogenous persistence after an effective treatment or from a "new exogenous acquisition of the bacteria" (11). Strains like E.coli O157:H7, O145, O26, O111 and O103 are the most common E.coli bacteria in America. Moreover, North American pulsed- field type 1 (NAP1/027/III) has been the most common C.diff strains that causes illness. However, since the emergence of C.diff Rybotype O78(10) in 2007 and E.coli O121 in 2013 there has been an increase interest and concern of the bacterium's

morbidity and mortality rates (12). Subsequently, following 2011 the trends and analyzing peak searches it is evident that newer strains have arose. This analysis supports the theory that Bacterial microorganism are capable of mutating, yielding new potent strains capable of causing epidemics (10).

We are in a post-epidemiologic transition era where we are now dealing with new emergent and/or drug resistant infectious diseases (10). In the past years, the pharmaceutical industry have shut down most of their Anti-Infective therapeutic units in order to focus on a more profitable investment that is the development of life long dependent chronic medications. Such actions put us all at great risk of infection. In a time where chronic diseases is the leading cause of death and worry it is important to not lose track on ever evolving microorganisms like E.coli and C.diff. Recently we have seen new strains of C.diff and E.coli bacteria, making them virulent and a cause for concern. Using real time data on search volume, evaluations were made on these bacteria in order to identify at-risk populations and help promote interventions.

The results of this study should be interpreted in the context of some limitations. First, given the sensitivity of Google Trends word search and diverse neologism based on levels of education there where inherent challenges as to what keywords that reflected the average online searcher to use. For the purpose of this study Keywords Cdiff and Ecoli were used without their proper nomenclature as a representation of the average online searcher. However, their appropriate nomenclature was also used for comparison purposes. The studies main focus was the community at large while comparing it to common word search behaviors of the scientific community. Secondly, Google trend results are subject to interpretation by the analyst. Third, our Google Trends evaluations were based on current syntactic possibilities and may change over time. Furthermore, these syntactic outcomes could potentially be subjected to non-representative sampling bias due to limitations of access of Google by the population, in which case cannot be discounted. The use of Google Trends has Spring Mix Blend | E. coli | CDC [Internet]. CDC. 2012 [cited 2015 Nov 14]. Available from:

proven to be extremely helpful in predicting future outbreaks while narrating it down to specific regions. Google Trends monitoring changes can potentially encourage public health officials to address such outbreaks and promote good health practices in infected regions.

REFERNCES

- Centers for Disease Control and Prevention. CDC - Clostridium difficile Infection - HAI [Internet]. 2015 [cited 2015 Nov 9]. Available from: http://www.cdc.gov/HAI/organisms/cdiff/Cdiff_infect.html
- Center for Disease Control and Prevention. CDC - Multistate Outbreak of Shiga toxin-producing Escherichia coli O121 Infections Linked to Farm Rich Brand Frozen Food Products [Internet]. 2013 [cited 2015 Nov 16]. Available from: <http://www.cdc.gov/ecoli/2013/O121-03-13/index.html>
- Center for Disease Control and Prevention. General Information | E.coli | CDC [Internet]. CDC. 2015 [cited 2015 Nov 14]. Available from: <http://www.cdc.gov/ecoli/general/index.html/>
- Center for Disease Control and Prevention. July 8, 2011: Outbreak of Shiga toxin-producing E. coli O104 (STEC O104:H4) Infections Associated with Travel to Germany | E. Coli [Internet]. CDC. 2011 [cited 2015 Nov 14]. Available from: <http://www.cdc.gov/ecoli/2011/travel-germany-7-8-11.html>
- Center for Disease Control and Prevention. March 23, 2012: Outbreak of E. coli O157:H7 Infections Linked to Romaine Lettuce | E. coli CDC [Internet]. 201AD [cited 2015 Nov 16]. Available from: <http://www.cdc.gov/ecoli/2011/romaine-lettuce-3-23-12.html>
- Center for Disease Control and Prevention. Multistate Outbreak of Shiga Toxin-producing Escherichia coli O157:H7 Infections Linked to Organic Spinach and Spring Mix Blend (Final Update) | Multistate Outbreak Linked to Organic Spinach and

<http://www.cdc.gov/ecoli/2012/O157H7-11-12/index.html>

- Goorhuis A, Bakker D, Corver J, Debast SB, Harmanus C, Notermans DW, et al. Emergence of *Clostridium difficile* Infection Due to a New Hypervirulent Strain, Polymerase Chain Reaction Ribotype 078. *Clin Infect Dis*. 2008 Nov 1;47(9):1162–70.
- Johnson S, Adelman A, Clabots CR, Peterson LR, Gerding DN. Recurrences of *Clostridium difficile* Diarrhea Not Caused by the Original Infecting Organism. *J Infect Dis*. 1989 Feb 1;159(2):340–3.
- L. Clifford McDonald, M.D., An Epidemic, Toxin Gene–Variant Strain of *Clostridium difficile*. *N Engl J Med*. 2005 Dec;353(23):2433–41.
- Lee H, Popodi E, Tang H, Foster PL. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci USA*. 2012 Oct 9;109(41):E2774–83.
- Responsible Reform For The Middle Class. The Patient Protection and Affordable Care Act [Internet]. United States Congress; 2010. Available from:<http://www.dpc.senate.gov/healthreformbill/healthbill04.pdf>
- Seifter A, Schwarzwald A, Geis K, Aucott J. The utility of “Google Trends” for epidemiological research: Lyme disease as an example. *Geospatial Health*. 4(2):13 5–7.