

**AN ANALYSIS OF CURRENT TOPICS RELATED TO HPV VACCINATION  
USING TWITTER AS A PUBLIC HEALTH TOOL**

By

**Priam Chakraborty**

B.S., Duquesne University, 2014

Submitted to the Graduate Faculty of  
Infectious Diseases and Microbiology  
Graduate School of Public Health in partial fulfillment  
of the requirements for the degree of  
Master of Public Health

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

**Priam Chakraborty**

It was defended on

**July 9<sup>th</sup>, 2015**

and approved by

**Thesis Advisor:**

Anthony J. Silvestre, PhD  
Professor, Infectious Diseases and Microbiology  
Graduate School of Public Health  
University of Pittsburgh

**Committee Member:**

Brian A. Primack, MD, PhD  
Associate Professor of Medicine, Pediatrics, and Clinical and Translational Science  
School of Medicine  
University of Pittsburgh

**Committee Member:**

Mackey R. Friedman, PhD  
Assistant Professor, Infectious Diseases and Microbiology  
Graduate School of Public Health  
University of Pittsburgh

Copyright by Priam Chakraborty  
2015

Anthony J. Silvestre, PhD

**AN ANALYSIS OF CURRENT TOPICS RELATED TO HPV VACCINATION  
USING TWITTER AS A PUBLIC HEALTH TOOL**

Priam Chakraborty, MPH

University of Pittsburgh, 2015

**ABSTRACT**

Human papillomavirus (HPV) is a sexually transmitted infection that can lead to many health complications including genital warts and certain types of cancer. Although HPV is a vaccine-preventable illness, many individuals continue to resist vaccination for both themselves and their children. The objective of this study was to use Twitter, a social media platform for microblogging, to assess views and attitudes towards HPV vaccination. Through a qualitative analysis of tweets posted by users, we hoped to gain a broad public opinion on the topic. Underlying objectives of this study included assessing the practicality of Twitter as a public health tool and determining the potential use of Twitter as a means to further the acceptance of the HPV vaccine.

After analyzing a random subsample of 2,000 HPV related tweets collected over a one-week time span, we found certain topics to be the center of discussion. The four categories that accounted for the largest proportion of tweets included news and media coverage of current events related to the HPV vaccine, the impact of receiving the vaccine on sexual behavior, and the safety and effectiveness of the vaccine. The public health significance of this research was to be able to use Twitter as an adjunct tool for identifying current reasons behind not vaccinating for HPV, and potentially for overcoming such barriers.

## Table of Contents

<b>1.0 Introduction</b> .....	<b>1</b>
<b>1.1 HPV infection and Gardasil</b> .....	<b>1</b>
<b>1.2 Problem Statement</b> .....	<b>2</b>
<b>1.3 Social Media and Public Health</b> .....	<b>3</b>
<b>1.4 Objectives of the Study</b> .....	<b>5</b>
<b>2.0 Review of Relevant Literature</b> .....	<b>6</b>
<b>2.1 Views among Young Adults</b> .....	<b>6</b>
<b>2.2 Views among Parents</b> .....	<b>7</b>
<b>3.0 Methods</b> .....	<b>10</b>
<b>3.1 Data Collection and Development of a Codebook</b> .....	<b>10</b>
<b>3.2 Codebook Development</b> .....	<b>12</b>
<b>3.3 Coding and Data Analysis</b> .....	<b>12</b>
<b>4.0 Results</b> .....	<b>13</b>
<b>4.1 Relevance</b> .....	<b>13</b>
<b>4.2 Population Sample</b> .....	<b>14</b>
<b>4.3 Inter-rater agreement</b> .....	<b>14</b>
<b>4.4 Sentiment</b> .....	<b>15</b>
<b>4.5 Safety and Effectiveness</b> .....	<b>16</b>
<b>4.6 Sexual Behavior</b> .....	<b>16</b>
<b>4.7 News and Media</b> .....	<b>17</b>
<b>4.8 Other Minor Categories</b> .....	<b>17</b>
<b>5.0 Discussion</b> .....	<b>18</b>
<b>5.1 Health reporting and promotion</b> .....	<b>19</b>
<b>5.2 Analysis of Sentiment</b> .....	<b>21</b>
<b>5.3 Minor Categories</b> .....	<b>22</b>
<b>6.0 Limitations</b> .....	<b>22</b>
<b>6.1 Twitter Data Collection</b> .....	<b>22</b>
<b>6.2 Complexity of Twitter Posts</b> .....	<b>23</b>
<b>6.3 Twitter Population</b> .....	<b>24</b>
<b>7.0 Future Directions</b> .....	<b>25</b>
<b>Appendix: Tables and Figures</b> .....	<b>26</b>
<b>Bibliography</b> .....	<b>33</b>

**List of Tables**

**Table 1 – Final Codebook .....26**

**Table 2 – Prevalence of Variables.....30**

## **List of Figures**

**Figure 1 – Number of Tweets from Unique Users.....31**

**Figure 2 – Binary Variables.....32**

## **1.0 Introduction**

### **1.1 HPV infection and Gardasil**

Human papillomavirus, commonly referred to as HPV, is a DNA papillomavirus that is transmitted through sexual contact and has been long known to infect humans, male and females alike. There are over 170 known strains of HPV, and while infections from these strains can often be initially asymptomatic, many strains have been shown to be linked to genital warts or cancer. (1) Persistent HPV infections are most commonly associated with cancers of the cervix, vulva, vagina, penis, and anus. (2) HPV has also recently been associated with other cancers, such as oropharyngeal cancer, in those who engage in oral sex. (3) In 1983, a study identified HPV 16 and 18 as causes of cervical cancer, and later studies have shown that these two strains are single handedly responsible for over 70% of all cases of cervical cancer. (4,5)

Fortunately, HPV is a vaccine preventable infection. The United States Food and Drug Administration approved Gardasil, a prophylactic vaccine that protects against four of the most prevalent types of HPV in 2006. By 2008, 41 states both approved and recommended Gardasil (6). The vaccine protects against HPV 6 and 11 that cause genital warts, and HPV 16 and 18, which as previously mentioned cause an overwhelming percentage of cervical cancer cases. (7) More recently, the same pharmaceutical company – Merck & Co. – established a newer version of the vaccine, which now protects against an additional five types of HPV for a total of nine strains of the virus. This new vaccine is called Gardasil 9, and was approved by the FDA in December 2014. (8)



## 1.2 Problem Statement

Unfortunately, even though HPV is a vaccine preventable infectious disease, and Gardasil has been shown to be safe and effective, vaccination rates remain low. The prevalence of HPV continues to increase in the population because of these low vaccination rates. In the United States, approximately 79 million people are estimated to be infected with HPV, and about 14 million are newly infected every year. (9) As expected, the health conditions commonly associated with HPV also have high incidence. About 36,000 sexually active people in the U.S. suffer from genital warts every year, and around 11,000 women develop cervical cancer every year. (9) These alarming statistics bring to question the reasons behind why, even after knowing the consequences of HPV, people choose to not vaccinate for the HPV. These reasons can be analyzed using a social ecological framework.

Gardasil is most effective if administered prior to sexual activity at a relatively young age (starting at 9 years old), but parents continue to not have their children vaccinated. This may be due to the stigma surrounding sexually transmitted diseases, and even more generally, the fact that discussing sexual health is taboo. (10) This is an interpersonal factor, as secrecy about sexual activity is often considered a social norm. Parents who conform to this norm, and are uncomfortable discussing intimacy issues, are less likely to be informed about the benefits of Gardasil, and thus less likely to have their children vaccinated. The decision could be individual level as well, as parents often times do not acknowledge the fact that their child may soon be or already be sexually active without them knowing. (11) For this reason, they do not recognize the need for having the child vaccinated for a sexually transmitted infection such as HPV.

Another problem may be that many individuals do not verify the reliability of a source before internalizing the information they receive. This is a complex factor, and may be

considered at the institutional, community, or interpersonal level. A religious institution may directly discourage vaccination – for example, if a family is a part of a cultural or religious community that has negative opinion of Gardasil, the parents may chose to not vaccinate their child. Similarly, if an influential other has heard “Gardasil scare stories” or has a negative view of the vaccine, they may persuade a parent away from having their child vaccinated. (12)

Lack of awareness or lack of access may also be a reason for not vaccinating. On an individual level, a parent’s socioeconomic status (SES) may hinder their access to care and therefore their access to vaccines. (13) More specifically, Gardasil is a vaccine that requires three doses, and therefore three visits to a provider. This could be expensive and improbable for some families. Institutionally, while many schools and colleges often require incoming students to be vaccinated for various infectious diseases, HPV is not one of them. Parents may be more likely to have their child vaccinated if an educational institution requires it. (14) Lastly, there is a lack of advocacy for the Gardasil vaccine in the health care system. PCP’s are not required to discuss HPV vaccination with patients, so parents are often left uniformed about the matter and do not consider having their children vaccinated.

### **1.3 Social Media and Public Health**

Social media has been characterized as one of fastest, most cost-effective, and interactive ways of spreading information and communicating with the public on a large scale. (15,16) More specifically, research has found that social media is commonly used as a source for the latest health information, particularly among teenagers and young adults ranging from 18 to 30 years old. It also suggests that this age group relies more heavily on the Internet for their health concerns than on their physicians. One example of a social media website being used in health

care can be seen in the way YouTube is used in dermatology. Many patients have used the media as a way to record and share the outcomes of different medications to aid other potential consumers. (17) Of all social media outlets, however, microblogging has gained much popularity in the past few years as a way to quickly exchange information and communicate.

Microblogging is defined as a medium that allows users to publicly share thoughts and opinions, solo images, or single videos in succinct posts. (18) To keep up with the fast paced life of the 21<sup>st</sup> century, microblogging encourages communicating in brief statements and has as a result become a preferred tool. While there a few different microblogging services out there, Twitter is the most popular. (19)

Twitter was created in 2006, and as of May 2015, it claims to have over 500 million users worldwide, of which 302 million are considered to be active users. (20) Members of this social media service include individuals of all age groups, communities, and backgrounds. Users use Twitter to express their opinions on current events, share new information with their followers and the general public, and communicate using short (140 characters) messages. Twitter allows users to get information out in a quick and efficient manner, and get almost instant feedback from the target audience. A wide range of businesses, including everything from supermarkets to airlines and banks, use Twitter to directly reach customer to both answer questions and concerns and take suggestions. In fact many public health agencies, such as the World Health Organization (WHO) and the Center for Disease Control and Prevention (CDC) and even on a local level such as NYC Department of Health, have adopted Twitter as well. (21) These agencies, among others, keep their followers updated with any vital health news. Academic and research institutions have recently started using Twitter as a research implement as well. (16)

Twitter has been a useful in providing real time updates and advice emergency preparedness operations during events such as Haiti's 2010 earthquake and Egypt's 2011 civilian uprising. (22) Earlier this year, researchers used Twitter to gain insight in to "vaping culture", also known as e-cigarette use, by collecting tweets on a day that was deemed to be "World Vape Day". (23) In fact, Twitter has also been successful in infectious disease monitoring and vaccine promotion as far back as 2009. During that year, researchers used Twitter for syndromic surveillance in which they tracked trends in complaints about flu like symptoms and compared it to the geographic location these tweets came from to monitor the spread of H1N1. (24) During the same outbreak, the health department of Alexandria, Virginia used tweets to promote flu vaccines as soon as they were available during the pandemic, and noted spikes in clinic visits almost immediately after. (25) Due to its diverse functionality, Twitter has recently gained popularity as a useful public health tool.

#### **1.4 Objectives of the Study**

As mentioned, there are many probable reasons behind low HPV vaccination rates, but most research done to assess views, barriers, and opinions on the issue have been done through methods such as interviews, surveys, and focus groups within small subpopulations. In a broadly diverse, constantly changing environment however, an all-inclusive view from a broader population could prove to be a useful tool. Due to the expansive reach of the network, we chose use Twitter to obtain a generalized public opinion on a popular public health issue.

The overarching goal of the study, therefore, was to use Twitter to assess attitudes and views verbally expressed about HPV vaccination and Gardasil. A qualitative analysis of the content obtained could then be compared to existing research on such issues to determine what

topics are currently pertinent to HPV vaccination and what the broad public opinion is regarding the matter. Our population of interest was the online Twitter community, and our study was designed to capture a fragment of data that would be representative of the global population that uses social media. Underlying objectives of this research include assessing whether Twitter is a useful tool in public health, and considering how to use data collected from this study in cooperation with already existing research to further the acceptance of the HPV vaccine.

## **2.0 Review of Relevant Literature**

Ever since the Gardasil has been FDA approved, much research has been done to assess views on this HPV vaccine. These assessments commonly included information on what participants know and do not know about the vaccine, and what factors individuals consider when deciding to be vaccinated or have their children vaccinated. Some addressed more specific concerns as well, such as access to care and cost of the vaccine. In terms of populations, some studies have focused on parental views on having their children vaccinated, while others have concentrated on vaccine acceptance among young adults for themselves. While most individual studies worked with small population sizes when used and analyzed collectively, they provided an overall inclusive view.

### **2.1 Views among Young Adults**

Knowledge about HPV and acceptance of Gardasil has been recurrently found to be high among young adults. (26-28) However, most studies that chose to focus on young adults recruited participant samples solely from college or university settings and typically used surveying methodology. A study done at a college in the United States Midwest in 2003 found

using questionnaires that, on average 74% of the young adults surveyed viewed the HPV vaccine positively and would endorse it. (26) A slightly different study, performed in 2001, but in the Northeast, made use of more specific questions and focused on knowledge about HPV and Gardasil rather than acceptance. This study found that on average, only 45% of the HPV related questions were answered correctly. (27)

A study done years later in 2006, recruited students from two distinct southeastern universities in the U.S. encountered comparable results. This study found that while 75% of their sample knew of HPV and associated vaccines, very few knew accurate or detailed information. Moreover, there were disparities within their sample: for example, females and older students were more accurately informed and showed a greater interest in the subject matter than others. (28)

While these findings show that many college students have some knowledge about HPV and the vaccine, the data indicated that the amount and extent of knowledge and acceptance has not significantly changed over time. Furthermore, there appears to be lack of information on non-college students that fall in to this age group. Although it is undoubtedly easier to recruit college students as it can be done on a campus, without information on the subpopulation of young adults who do not attend college, it is not possible to fully assess knowledge, views, and attitudes on HPV and Gardasil among adolescents using these methods.

## **2.2 Views among Parents**

Overall, many researchers who worked specifically with parents have found that parents are more knowledgeable of HPV and more of accepting of the Gardasil shot than generally thought to be. Although the percentage of parents who would be willing to vaccinate their child

and endorse the vaccine varied greatly between studies, almost all research done on the topic found that the majority of parents (over 50% of those recruited) had a working knowledge of HPV and supported the vaccine. Various research methods were used to contact parents for such studies. One report from 2004 used a detailed survey method, and found that 60% of the parents who participated had a general understanding of HPV and endorsed the vaccine. (29) Another study done in 2007 that conducted random digit-dialing (RDD) interviews and exclusively interviewed parents who had daughters found that 75% of those questioned would be willing to vaccinate their daughter before 13 years of age. (30)

Other research projects have focused on comparing vaccine acceptance between parents and non-parents, and within the subset of parents comparing between choosing to vaccinate one's self versus choosing to vaccinate a child. One questionnaire-based study performed in 2006 did notice differences; researchers found that while 77% of women would be willing to get the vaccine themselves, only 67% would choose to have their daughter vaccinated and 66%, would opt to have their son vaccinated. (31) A different research group that questioned both young adults who were parents and those who were not, found that the acceptance of HPV immunization was roughly consistent when it came to vaccinating themselves and vaccinating their children. (32)

Although HPV knowledge and Gardasil acceptance rates were generally high in most of the studies conducted, there were a few concerns that repeatedly came up among those who were surveyed. The study that noted the highest vaccine acceptance rate of 75% (30) also noted that concern regarding an increased likelihood of their child engaging in sexual activity was the most predominant reason that kept the remaining 25% against or undecided about the vaccine. Other studies have found that the most important factors in Gardasil acceptance were related to

concerns regarding vaccine safety along with concerns as to whether this vaccine was actually necessary and how much the vaccine would cost. (26,29)

Many of the aforementioned reports also chose to ask questions about what could possibly change a participants' decision to vaccinate or not vaccinate. Some of the topics that were at the top of this list were physician recommendations and the general use and acceptance of the vaccine on both a local and global level. In other words, parents were more likely to endorse the Gardasil vaccine if it was either being widely being used or being promoted by their primary care providers. Both of these influences indicate that decisions were more likely to be made based on personal or peer influences rather than facts alone. A group of researchers, who measured HPV knowledge and vaccine acceptance both before and after conducting an informational intervention, reported similar findings in 2006. Even after being presented with research results and CDC endorsed facts, most participants stuck to their original decision of whether or not they chose to endorse the vaccine. (33)

Overall, while these previous studies provide some insight in to what certain populations think about HPV vaccination, and why they hold their views, there are some perceived benefits to using an online data collection method over using questionnaires and interviews. One such benefit is being able to encompass a broader and more diverse population. Twitter is a globally accessed social media website, and allows us to get varying viewpoints from all over the world. Furthermore, people are likely to be more expressive when using an online interface. For example, while someone may be reluctant to answer in depth questions when directly interviewed or propositioned with a questionnaire, they may be willing to share more information online where they are not required to share their identity fully. For these reasons,



Twitter was used in our study to monitor current trends and attitudes towards HPV and HPV vaccination.

## **3.0 Methods**

### **3.1 Data Collection and Development of a Codebook**

The program used to screen and collect live data from Twitter was Twitter's *Public Streams* Application Programming Interface (API). (34) Python programming interface, using *Python(x,y)* software (35), was used to write a customized code that was built on the *Twython* package. (36) This customized code allowed us to selectively retrieve data from the Twitter API that was particularly relevant to our topic. When Twitter's public streams were used in previous studies to extract a random sample of tweets from the Twitter *firehose* – the full stream of all Twitter data – there were often technical difficulties in the program when the stream flow collected exceeded 1% of the entire *firehose* flow. (37) However, because our topic was highly specific and the code created to filter the tweets used very definitive terms, the stream did not exceed the 1% threshold. We were able to successfully elicit data we were interested in, and no known relevant content was left out.

IRB approval was obtained by the University of Pittsburgh for permission to use the tweets collected from human subjects for research. A test run was performed in November 2014, and HPV and Gardasil related tweets were collected for a brief time period. This information was used to decipher what keyword filters would be the most fitting for the study. Certain terms were added during the process while others were taken away in an effort to create a list that would preserve all pertinent data but also fit the 1% threshold. For example, the word “cervical” on its own was initially on the list. It was later removed because many of the tweets collected

containing the word were related to cervical spine and neck problems instead of HPV or cervical cancer. The following is the finalized list of keywords that were employed during data collection: *HPV, papilloma, pappiloma, papiloma, pappilomavirus, gardasil, gardisil, guardasil, guardisil, cervarix, cervical shot, cervical shots, cervical vaccine, cervical vaccines, cervical vax, cervical vaxine, cervical vaxines, cervical vaxx, cervical vaxxine, cervical vaxxines, cervical vaccination, cervical vaccinations.*

In February 2015, the aforementioned list was used to collect data from 1:45 p.m. (GMT-5; Eastern Time, U.S.) on Friday, February 6<sup>th</sup>, 2015 till 3:11 p.m. on Thursday, February 26<sup>th</sup>, 2015. While HPV related tweets were available during this entire given time frame, it was decided that we would solely use tweets from 12 a.m. on Saturday, February 7<sup>th</sup>, 2015 till 11:59 p.m. on Friday, February 13<sup>th</sup>, 2015. This time frame was selected for a few reasons. It provides exactly one set of sample tweets from each day of the week, eliminating that as a confounding variable. Furthermore, it excludes holidays, such as Valentine's Day, which may skew the nature or quantity of the tweets away from the normal. Lastly, while there were some short lapses in proper elicitation during the twenty days where tweets were unable to be retrieved for a brief time period – the chosen week was void of any such lapses. This confirms that all HPV and Gardasil tweets during the one-week span were included in the study. This process resulted in a total of 20,408 usable tweets. A random sample of 2,000 of these tweets was then selected for coding. In an effort to maintain IRB specifications, all personal identifiers such as Twitter usernames were left out during the coding process.

### **3.2 Codebook Development**

A primary codebook was developed using a general framework created using literature on the topic and the set of sample tweets collected in November 2014. It included categories related to or commonly associated with HPV vaccination such as: *positive-* and *negative sentiment, access to vaccination, policy-related, cost-related, and related to the new 9-valent vaccine*. The broad original framework was then systematically altered and made more specific using a grounded theory approach in which qualitative data was analyzed to identify recurring themes and ideas.

From the total of 20,408 tweets that were collected, smaller subsets were then used to analyze relevance by two independent coders. Sets of 200 tweets were used at a time. Both coders identified any disagreements, conferred with supervising researchers for feedback, and then worked towards modifying definitions for more precise codes. Inter-coder reliability tests were performed using SPSS for all of the original constructs to identify areas that required redefinition or finer detail. After three sets of collaborative coding, a final codebook was decided on using both the original theoretical information and grounded constructs from the collected data. All codes were binary, and therefore the codebook consists of nine dichotomous constructs. The codebook also lists any associated definitions, and provides examples of what does and does not fit in to a given category. The final version is attached (Table 1).

### **3.3 Coding and Data Analysis**

The subset of 2,000 tweets used for analysis accounts for approximately 9.8% of the total sample of 20,408 that was collected. Of this subset, both coders coded three distinct sets of 200 tweets, for a total of 600 tweets. Percent agreement and Cohen's Kappa statistics were calculated

to establish inter-rater reliability for contextual variables. After broadly coding the sample using the binary variables, the two coders were able to attain a desirable level of agreement and establish a final codebook. The remaining 1,400 tweets were then coded by one of the two coders using a grounded theory framework and guidance from a senior researcher with qualitative analysis experience. The re-defined, final codebook was also used to re-code any of the 2,000 tweets that were coded prior to the establishment of a final codebook. Microsoft Access software was used to organize and code all collected data.

From the subsample of 2,000 tweets, 1,887 tweets or 94.35% were found to be directly relevant to HPV, and approximately 113, or 5.65% were found to be irrelevant. These 1,887 tweets were then further analyzed using qualitative methods to identify common trends or ideas listed in the codebook or otherwise. For example, was the twitter population that was talking about HPV more concerned about the safety of the Gardasil vaccine or the efficacy of the vaccine? Were the tweets coded to fall in one of these two categories also typically coded to be in the other as well? SPSS was used for all statistical analysis that was performed after coding was completed.

## **4.0 Results**

### **4.1 Relevance**

As mentioned, 94.35% were found to contain content directly relevant to HPV vaccination or Gardasil. The remaining tweets were excluded for further analysis. There were two common reasons for excluding a tweet due to irrelevance. One reason was that tweets received through the stream were sometimes cut short due to the 140-character limit enforced by Twitter. When users “re-tweeted” information or links posted by other users, the full length of

the original message would be cut short leaving coders unable to fully confirm whether or not the tweet was pertinent to HPV or Gardasil. The second issue was tweets that had relevant words, but the words were being used in a non-related context. For example, if HPV was used in a tweet but as an acronym for something other than human papillomavirus, it was not considered relevant.

## **4.2 Population Sample**

In order to assess whether or not our sample was representative of a broad population, we calculated how many of our relevant tweets were coming from unique users. We found that the majority of our tweets, 88.39%, came from unique users – this means that of our working sample of 1,887 tweets, 1,668 tweets were from distinct users. A small percentage, 9.96%, of the tweets came from users who posted between two to five times. An even smaller percentage of the tweets, 1.64%, represented tweets from users who accounted for more than five of the tweets in our sample. The highest frequency of tweets from a single user was found to be twelve – only one user fell in to this category. Overall, our sample consisted mostly of tweets from distinct users. (Figure 1)

## **4.3 Inter-rater agreement**

Inter-rater agreement was calculated during codebook development for three distinct sets of 200 tweets. During the first round of coding, Kappa scores varied greatly ranging from -0.0139 to 1, and therefore indicated disagreement between coders. After codebook modification, Kappa values were calculated for the second set of coded tweets and fell between 0 and 1. The codebook was altered one last time (Table 1), and all Kappa scores for the third set of tweets fell

between 0.5690 and 1. Coding disagreements were typically related to identifying human emotions such as sentiment. However, as suggested by Fleiss, Kappa values falling between 0.41 and 1 still indicated moderate to almost perfect agreement. (38) Furthermore, the results were in-line with those reported in recent studies using human coders and machine learning algorithms to characterize Twitter content. (39-42) The codebook was not modified further and the remaining tweets were coded according to the finalized variables and definitions.

#### **4.4 Sentiment**

A total of 98 of the 1,887 relevant tweets, approximately 5.19%, were marked to have positive sentiment. In order to identify positive sentiment, coders looked for terms that encouraged vaccination or described it in a positive manner. This included but was not limited to words and phrases such as *works well*, *recommend*, *vaccines work*, and *vaccinate your kids*. About the same number of negative sentiment tweets were collected from our relevant tweet subset (95 tweets, or 5.03%), indicating that both sentiments were equally expressed in our data. Coders looked for terms that discouraged Gardasil vaccination or expressed a negative view towards it. Words identified included but was not limited to *beware*, *destroys lives*, *hurt*, and *mystery illness*. While sentiments were sometimes directly incorporated in to the tweets, we found that they were also sometimes a part of the hashtags at the end of the tweet. For example, #CDCWhistleBlower was a common hashtag used along with negative sentiment tweets and seemed to be targeted responses to tweets from the CDC account that encouraged vaccination.

## 4.5 Safety and Effectiveness

The safety and effectiveness of the Gardasil vaccine was initially under one coding category. However, after the first round of coding, coders found that tweets that commented on one of these two factors did not necessarily comment on the other. For this reason, safety and effectiveness of the vaccine were coded for separately during the second round of coding, and any tweets from the first set were recoded accordingly. A total of 210 tweets, or 11.13%, commented on the safety of the vaccine (Table 2). Of these, 38.1% implied Gardasil to be unsafe, while 61.9% of them deemed the vaccine safe (Figure 2). Examples of phrases that implied safety include *'it is safe'* and *'safety of the HPV vaccination is reaffirmed'*. Examples of phrases that were marked as tweets saying the vaccine is unsafe include *'Gardasil ruins live'* and *'girl dies shortly after receiving HPV vaccine'*. Of all the tweets mentioning safety, there were about 1.62 times more tweets stating Gardasil was safe rather than unsafe.

## 4.6 Sexual Behavior

Another prominent category that surfaced after the first set coding was Gardasil's influence of sexual behavior. A total of 533 tweets, or 28.25%, contained content concerning Gardasil's effect on sexual behavior (Table 2). Interestingly, an overwhelming percentage (96.81%) of these tweets supported the idea that receiving the vaccine either does not influence or does not increase sexual activity or risky sexual behavior (Figure 2). That is 30.35 times more than the percentage of tweets (3.19%) that accused Gardasil of increasing risky sexual behavior. A large number of these tweets seemed to reference a newly published (February, 2015) article from Harvard Medical School titled "HPV Vaccination not Linked to Riskier Sex". (43) As a result, the tweets coded to indicate no influence on sexual behavior contained expressions like

*'HPV vaccine will not turn your daughter in to a slut'* or *'HPV vaccine linked to less risky behavior'*. The tweets branded as accusing the vaccine of increasing risky sexual behavior contained statements such as *'HPV Vaccines make you promiscuous'*. (Table 2)

#### **4.7 News and Media**

We also found that many of the 1,887 tweets were either about something published or reported recently about Gardasil, or direct re-tweets of posts from major newspapers, magazines, or TV channels that had to do with Gardasil. A total of 787, 41.71%, tweets were coded as having content related to news and media (Table 2). Coders identified such tweets with both key terms such as *coverage*, *article*, and *story* and direct mentions of known media corporations such as *@TorontoStar*, *@USATODAY*, or *@ABC*. (Table 1)

Upon closer analysis, we found that many of the news and media tweets seemed to be about a recent HPV vaccine scare story that was published and then retracted in the Canadian newspaper *The Toronto Star* (44), with people expressing personal opinions about the article such as *'Never lost respect for a publication as fast as I lost respect for @TorontoStar with their HPV vaccine coverage'*. Many of the direct re-tweets seemed to be from different publications that were covering the aforementioned Harvard Medical School study about Gardasil and sexual behavior – *'HPV vaccines do not lead teen girls to risky sex. Via @USATODAY'* is one such example.

#### **4.8 Other Minor Categories**

The codebook contained four other categories – legal and policy matters, barriers to vaccination, the Gardasil9 vaccine, and parental attitude – but a very small percentage of our



sample fell in to these four categories. Key terms used to identify legal or policy related tweets included *conservative* or *liberal*, or direct mention of politicians or government agencies such as ‘Governor Perry’s Gardasil vaccine mandate cost young girls lives’. Approximately 1.43% of the tweets were a part of this group. Any posts discussing *access to care*, *cost*, etc. were considered to be related to barriers to vaccination – 0.90% were a part of this faction. Key words such as *9-valent* or *new vaccine* were used to identify tweets regarding the newly FDA approved Gardasil9 – only 0.58% of tweets were noted in this category. Lastly, any posts that used terms such as ‘*my child*’ or ‘*my son/daughter*’ were marked as parental attitude – only six tweets (0.32%) were in this group. (Table 2)

During data analysis, coders noted a few topics had a fairly strong presence in our data set but were not listed as variables within the codebook, and therefore not analyzed further. The biggest of these variables was marketing for a new HPV medication. There seemed to be a stream of tweets posted to advertise Gene-Eden-VIR, a natural antiviral that fights HPV infection. (45) Examples of some other topics that were recurrent include disparities in HPV infection rates according to race, gender, and sexual orientation, and HPV screening discrepancies across the globe. However, because these topics do not provide direct insight in to the disparities that exist in HPV vaccination rates, they were not used for further analysis

## **5.0 Discussion**

Through one week of continuous data collection using *Twitter’s Public Stream API*, we were able to use Twitter as a public health tool to gain insight into current issues, opinions, and concerns regarding HPV vaccination. While some of the posts collected were on topics we expected to encounter, such as the newly FDA approved 9-valent Gardasil vaccine and opinions

on current HPV vaccination policies, a large portion were about matters we didn't anticipate to be in the spotlight. Issues in modern day health reporting and newly published scientific studies on the effects of HPV vaccination on sexual behavior are two such examples. Through analysis of tweets, we were also able to assess the usefulness of Twitter as an adjunct method in evaluating opinions on and possible barriers to HPV vaccination, and more generally the applicability of Twitter as a qualitative research instrument in public health.

### **5.1 Health reporting and promotion**

“News and media” seemed to be the variable that singularly accounted for the largest proportion of tweets with 787 out of 1,887 or 41.71% of the posts falling under this categorization. These tweets mainly concerned one of two topics – the first of these being the importance of proper health reporting in regards to Gardasil. One particular article that appeared to be the center of discussion was a front-page article from Canadian newspaper, *The Toronto Star* titled “A wonder drug’s dark side”. (44) The article was published in February 2015, shortly prior to our week of data collection. The piece was said to be an anecdotal Gardasil scare story that was later found to lack sufficient scientific accuracy, and was therefore heavily criticized by the medical, scientific, and public health communities. *The Toronto Star*, colloquially known as *The Star*, acknowledged this criticism and resumed to retract the article and replace it with one that reckoned Gardasil to be safe. The piece was titled, “Science Shows HPV Vaccine has no Dark Side.” (46)

The overwhelming public health response that was acknowledged by the paper was also evident in the tweets in our sample. Many tweets collected on the topic criticized the paper through statements such as “@TorontoStar botched a story about #HPV vaccine,” and “this is

appalling, ignorant, irresponsible journalism”. Moreover, twitter also seemed to serve as a platform for credible sources to rectify misconceptions about Gardasil – one tweet directly stated that the “response may not have been as vocal without docs on social media”. One prominent name is Dr. Jen Gunter, a physician and health blogger who promptly reacted to The Star’s piece to correct the misinformation it presented. (47) There were many tweets that stood in her support by posting statements like Listen to “@DrJenGunter take down that atrocious @TorontoStar story on Gardasil.” This finding is in concordance with the past studies that found physician recommendation having a strong influence on the public’s opinion of the HPV vaccine. (18) As a result of the controversy that surrounded this health-reporting incident, numerous of the tweets that were about *The Toronto Star* article were also about vaccine safety (11.13%) or effectiveness (5.94%). Additionally, of those who judged Gardasil to be safe, many referenced either *The Star*’s remedying article or posts from public health advocates who stepped up to defend the vaccine. “‘The HPV Vaccine has No Dark Side.’ Thank you to the 67 doctors & researchers who signed this @TorontoStar editorial” – is one such example.

The second topic that kept resurfacing in our news and media tweets was the Harvard Medical School study titled – *HPV Vaccination not Linked to Riskier Sex*. (43) Researchers at these institutions used health insurance records to monitor STI prevalence in girls who received the HPV vaccine and compared the rate to girls who did not. As can evident by the title, they found that receiving the vaccine was not associated with unsafe sexual practices. An abundant number of tweets acknowledged the lack of association (27.34%) therefore made statements like “No, the HPV vaccine will not drive teens to have wild, promiscuous sex” in support of the newly published data. What we also noticed was that study was covered by many influential newspapers and media networks, and then retweeted by users who followed these agencies on

Twitter. Two examples are – “@latimeshealth : For teen girls, getting older is a risk factor for STIs. HPV vaccine isn't” and “Teen HPV vaccine does not spur riskier sex. (Reuters Health) – to counter the fears of some parents.” These tweets also fell in to both the “news and media” and “the sexual health” categories.

Responses surrounding both the Gardasil news story on *The Star* and the medical study on Gardasil's influence on sexual behavior were amplified using Twitter as a platform. *The Star* is a reputable newspaper, and by publishing a misleading story on HPV vaccination, they ran the risk of swaying the public away from a drug that can serve to reduce the cervical cancer rates in our community. In this case, Twitter served as a medium through which clinician's were able to answer questions users had and provide facts that proved Gardasil to be safe. Harvard Medical School found that statistics show no association between receiving the HPV vaccine and engaging in risky sexual behavior. Twitter similarly served as a channel through which a reputable academic institution was able to share valuable facts with the general public that could work towards encouraging HPV vaccination in the community. These two examples go to show how Twitter could be used as a platform for the diffusion of proper health reporting and health promotion. More generally, the findings demonstrate the impact of Twitter has as a public health tool.

## **5.2 Analysis of Sentiment**

The analysis of sentiment on HPV vaccination proved to be a harder task than anticipated. As mentioned, only tweets that very directly promoted or discouraged Gardasil were labeled as positive or negative. This methodology lead to an almost equal percentage of tweets falling under each of the two sentiments. For example, tweets that stated things “#HPV vax helps

prevent HPV types that cause most cervical #cancer #themoreyouknow,” may likely have been coming from individuals who carried a positive sentiment about the vaccine but did not explicitly encourage vaccination. Therefore, while our coding and analysis show equal representation of positive and negative sentiment, this may not be an accurate depiction of feelings our target population had towards the vaccine.

### **5.3 Minor Categories**

There are some topics that we anticipated more posts on, and therefore incorporated them as variables in the codebook. One such category was the 9-valent Gardasil vaccine that was approved by FDA less than a year ago in November, 2014. However, a mere 0.58% of our sample tweets were about Gardasil9. There are a few things this could imply including the possibility that the Twitter community is either predominantly uninformed about the new vaccine, or does not take interest in the matter. Some other areas that fell in to this category include legal and policy matters, barriers to vaccination, and parental attitude. Individually, these variables represented only 0.32-1.43% of all tweets. We were unable to draw any conclusion on these subject matters due to the limited amount of data we were able to collect.

## **6.0 Limitations**

### **6.1 Twitter Data Collection**

One factor that limited the generalizability of this study was the relatively short one-week span over which tweets were collected. While we were able to collect valuable information on a few specific topics related to HPV vaccination, including opinions on current Gardasil coverage in newspapers and newly published data on the effect of Gardasil on sexual behavior, these

topics are only representative of issues that were important that particular week. Additionally, prior to starting Twitter data collection, we were not able to pick or even gauge what topics would be the focus of our research. On social media, new issues surface regularly, but get buried just as quickly when a new matter commands the public's attention. For this reason, information collected and analyzed from one week of Twitter posts has limited applicability on its own. Furthermore, while using 2,000 randomly selected tweets made analysis of the data manageable, it may have hindered an all-encompassing examination of the tweets collected during that time frame.

## **6.2 Complexity of Twitter Posts**

Another significant limiting factor was using a strictly structured coding system to interpret the language present in tweets. While a well-structured codebook allowed us to approach qualitative data with a certain level of consistency, it also restricted the incorporation of some natural human components into analysis. Sarcasm, for example, was difficult to identify with certitude and was therefore factored out – all tweets were taken plainly, for face value. This also limited the coders' ability to identify sentiment. Only very direct statements either encouraging or discouraging HPV vaccination were coded as containing sentiment. We also refrained from going to or using any of the outside links users included in their tweets. Our intention was to focus on statements directly made to analyze personal views and opinions. While this aided in creating an organized framework, it retracted from our ability to check for factuality in the tweets we collected.

### **6.3 Twitter Population**

The final factor that limited the generalizability of our data was our population itself. Twitter is a platform in which all posts are voluntary. Because of this, those who have generally strong opinions about a subject matter are more likely to tweet about it. Therefore, the views and opinions we collected on Twitter about HPV vaccination may not represent the view of the general public but rather the views of those who are robustly either for or against HPV vaccination. Likewise, different big agencies or public figures may be overrepresented since such accounts have a generally large number of followers and thus a larger number of re-tweets. Lastly, due to the nature of social media, it is difficult and sometimes not possible to determine accurate demographic information (i.e. age, gender) on the population we collect data from.

An underlying objective of this research was to determine the practicality of Twitter in public health research. Although there are drawbacks to using Twitter as means to assess views related to a certain topic, there are some advantages as well. Primarily, collecting data through Twitter is fast and efficient; it requires less time and labor than methods such as interviews and surveys. Additionally, Twitter is a real-time device. Due to the speed at which data can be accessed and collected, Twitter allows researchers to gain insight in to the most current matters. Surveys or interview questions are generally made ahead of time, running the risk of becoming outdated and targeting issues that are no longer a top priority. Lastly, Twitter fosters instant feedback and communication. If researchers desired to, they can use Twitter to directly reply to users with questions or concerns.

## 7.0 Future Directions

Twitter, and more generally social media, has over the years become incorporated in to the everyday lives of many individuals – our research shows that public health researchers and workers could use this to their advantage. Twitter is already being used as a health promotion tool at the individual level. This was evident when doctors who use Twitter rectified the misinformation on Gardasil that was published on *The Toronto Star*. Universities are also using Twitter as a means to communicate new findings as was seen by the vast number of tweets in our data that shared information from the Harvard Medical School study about HPV vaccination and promiscuity. These results imply that Twitter can be used as a way for public health agencies and advocates to spread factual information about Gardasil to work towards increasing HPV vaccination. Moreover, because our research focused on one week of data, we collected information on a narrow range of topics related to HPV vaccination. One way to advance our findings would be collect similar data longitudinally for a longer period of time to monitor changes in topics and opinions. This would open windows for public health workers to address issues as they come, and even use twitter as a way to communicate with the public and answer questions they have on the HPV vaccine. Overall, due to Twitter’s wide functionality, continued research and use of Twitter as a public health tool has the potential to address issues currently preventing optimal rates in HPV vaccination.



## Appendix: Tables and Figures

Table 1 – Final Codebook

VARIABLE	TYPE	DESCRIPTION	EXAMPLES
Relevant	Binary	<p>Does the tweet contain a reference to HPV or HPV vaccination? Mark yes if tweet includes keywords such as “Gardasil” “HPV” “cervical cancer”.</p> <p>0 – No 1 – Yes</p> <p>**hashtags count (i.e. #Gardasil #HPV)</p>	<p>0-“seemed as if the cervical vertebrae were filled with fluid iron.”</p> <p>1-“Gardasil researcher is against the vaccine another myth debunked <a href="http://t.co/Ee0sBSgQ1w">http://t.co/Ee0sBSgQ1w</a>”</p> <p>1-“@dkegel @FDAWomen It's absurd. "Would you like a regular or supersized #hpv vaccine?" Absurd.”</p>
Positive Sentiment	Binary	<p>Are there any positive keywords used to describe the HPV vaccine/Gardasil? Are people encouraging or promoting the vaccine?</p> <p>Keywords include but are not restricted to: <b>vaccinate, recommend, great, works well, kills the HPV virus, awesome</b></p> <p>0 – No 1 – Yes</p> <p>*hashtags count ( i.e. #VaccinateYourKids, #VaccinesWork) **If there is a strong positive word used in the tweet that is NOT included in our list of keywords, mark as positive sentiment AND mark as important. ***If tweet simply states that the vaccine is safe or effective, <b>do not</b> categorize as positive sentiment.</p>	<p>0-“Now she's prob pregnant and you're a carrier of HPV.”</p> <p>1-“@undefined Get the HPV vaccine. It is safe [comma] effective [comma] and recommended for girls 11 and 12 years of age with a catch-up through age 26...”</p> <p>1-“Two genotypes of #HPV cause and estimated 70% of all #cervical #cancers. #vaccinate against HPV to reduce your risk.*”</p>
Negative Sentiment	Binary	<p>Are there any negative keywords used to describe</p>	<p>0-“Wonder if you told her about</p>

Table 1 Continued		<p>the HPV vaccine/Gardasil? Is there skepticism about the safety of the vaccine?</p> <p>Keywords include but are not restricted to: <b>hate, complaints, victim, devastation, regret, sad, died, problem, pain, hurt infertility, destroys lives, lies, propaganda, tricked, criminal, kill, mystery illness, plaguing, beware, bad</b></p> <p>*hashtags count ( i.e. #CDCWhistleBlower)</p> <p>**If there is a strong negative word used in the tweet that is NOT included in our list of keywords, mark as negative sentiment AND mark as important.</p> <p>0 – No 1 – Yes</p>	<p>the HPV. I kinda doubt it.”</p> <p>1-“ #ladygaga Mystery Illness Linked to Gardasil <a href="http://t.co/PK6jwwtaZL">http://t.co/PK6jwwtaZL</a>”</p> <p>1-“RT @karienne_: I hate doctors . you gave me a gardasil shot and I'm telling you I had a bad reaction but you blow it off cos medicine isn't”</p> <p>1-“RT @makerwithin: Gardasil Vaccine: One More Girl Dead   Health Impact News <a href="http://t.co/XIxRqu67zy">http://t.co/XIxRqu67zy</a> “</p>
Barriers to Vaccination	Binary	<p>Is the tweet related to the barriers to being vaccinated for HPV? Some known examples of barriers are access to care and cost of vaccination.</p> <p>Keywords include but are not restricted to: <b>too expensive, unavailable, barriers, low screening, access to care, cost-effective</b></p> <p>0 – No 1 – Yes</p>	<p>1-“Tel Aviv startup wants to bring cervical cancer assessment to women who have access to cellphones but not to doctors <a href="http://t.co/XOwZqbc6Tt">http://t.co/XOwZqbc6Tt</a>”</p> <p>1- “Barriers to HPV vaccination: MedlinePlus Health News Video.”</p> <p>1- “Every parent should read. HPV vaccination programs have not been shown to be cost effective”</p>
Parental Attitude	Binary	<p>Is the tweet related to opinions, views, behaviors, or attitudes parents have related to the HPV vaccine?</p> <p>Keywords include but are not restricted to: <b>my kid/ son/ daughter/ child</b></p> <p>*For tweets specifically from parents</p>	<p>0 – “Protect Your Daughters from Cervical Cancer [RETURN] [RETURN] HPV vaccines are given as a series of three shots over 6 months to... <a href="http://t.co/55q9gHR4C5">http://t.co/55q9gHR4C5</a>”</p> <p>1- “HOW COME NOW..... THEY RECOMMEND THE HPV</p>

Table 1 Continued		0 – No 1 – Yes	VACCINE FOR BOYS?? Last time I checked MY SON HAD NO OVARIES! [RETURN] #CDCwhistleblower #vaccines”
Legal Policy and Government	Binary	Is the tweet related to public or private regulation of Gardasil and HPV vaccination?  Key words include but are not restricted to: <b>public, regulation, ban, allowed, approve (d), FDA, medical (legal status), conservatives, liberals, right wing, left wing</b>  *Tweets about politicians and their stance on HPV vaccination also fall under this category  0 – No 1 – Yes	0-“RT @FDAWomen: It is NOT known how much condoms protect against #HPV. Areas not covered by a condom can be exposed to the virus: <a href="http://t.co/â€œ">http://t.co/â€œ</a> ”  1-“RT @VaxCalc: Utah health official bans Gardasil [comma] says Merck exaggerated benefits and FDA approved too quickly <a href="http://t.co/OFYX0sftqf">http://t.co/OFYX0sftqf</a> #CDCvâ€œ”  1- “Governor Perry's Gardasil vaccine mandate cost young girls lives. [break] <a href="http://t.co/F5Tcmo8rY6">http://t.co/F5Tcmo8rY6</a> ”
Vaccine safety	Binary	Does the tweet comment on the whether or not the HPV vaccine is safe?  Keywords include but are not restricted to: <b>safe, unsafe, ruins lives, dead, ruins lives, dead, dangerous</b>  0 – vaccine is unsafe 1 – vaccine is safe  *Hashtags count (i.e. #ruinslives, #vaccineswork)	0 – “Healthy 12-year-old girl dies shortly after receiving HPV vaccine <a href="http://t.co/E87qpOzRZX">http://t.co/E87qpOzRZX</a> ”  1-“@undefined Get the HPV vaccine. It is safe [comma] effective [comma] and recommended for girls 11 and 12 years of age with a catch-up through age 26...”  1 – “@drjennyblake: Safety of HPV vaccination is reaffirmed- we can all prevent cancer. <a href="http://t.co/4kSuocfu4F">http://t.co/4kSuocfu4F</a> ”
Vaccine effectiveness	Binary	Does the tweet comment on whether the HPV	1 – “@drjennyblake: Safety of

Table 1 Continued		<p>vaccine works/ is effective?</p> <p>Keywords include but are not restricted to: <b>works, doesn't work, protects, effective, prevents</b></p> <p>0 – vaccine is ineffective 1 – vaccine is effective</p> <p>*Hashtags count (i.e. #GardasilWorks, #PreventCancer)</p>	<p>HPV vaccination is reaffirmed- we can all prevent cancer. <a href="http://t.co/4kSuocfu4F">http://t.co/4kSuocfu4F</a>”</p> <p>1 – “@undefined A recent study by the CDC showed that the HPV vaccine is very effective and helped to lower HPV infection rates in teen girls by”</p>
9-Valent Vaccine	Binary	<p>Is the tweet about the movement from the 4-valent vaccine to the 9-valent vaccine?</p> <p>Key words include <b>new vaccine, updated drug, 9-valent</b></p> <p>0 – No 1 – Yes</p>	<p>1- “Merck announced that it’s<sup>a</sup> investigating a 9-valent HPV vaccine that protects against nine total types of HPV <a href="http://t.co/DkKAW13e85">http://t.co/DkKAW13e85</a>”</p>
Sexual Behavior	Binary	<p>Is the tweet related to the effect (or lack thereof) of HPV vaccination has on sexual behavior, habits, or diseases?</p> <p>Key words include but are not restricted to: <b>slut, promiscuous, unprotected sex, (less/more) risky behavior, STIs</b></p> <p>0 – vaccine does NOT lead to increased risky sexual behavior <b>OR</b> vaccine leads to decreased sexual behavior 1 – vaccine leads to increased risky sexual behavior</p>	<p>0 - “The Craziest Anti-Vaxx Argument: HPV Vaccines Make You Promiscuous - Daily Beast <a href="http://t.co/4x9nPBHD0S">http://t.co/4x9nPBHD0S</a>”</p> <p>0 - “HPV vaccine linked to less-risky behavior: By Roxanne Nelson Contrary to concerns that getting vaccinated against[...] <a href="http://t.co/RAoO7m8YI5">http://t.co/RAoO7m8YI5</a>”</p>
News/Media	Binary	<p>Is the tweet about the media or news coverage related to HPV vaccination/Gardasil?</p> <p>Key words include <b>journalism, health reporting, coverage, article, story, news</b></p> <p>*Any tweets from newspapers or magazines on</p>	<p>1 – “RT @onceuponA: I have never lost respect for a publication faster than I lost respect for @TorontoStar with their HPV vaccine coverage.”</p> <p>1- “RT @RCScience: HPV</p>

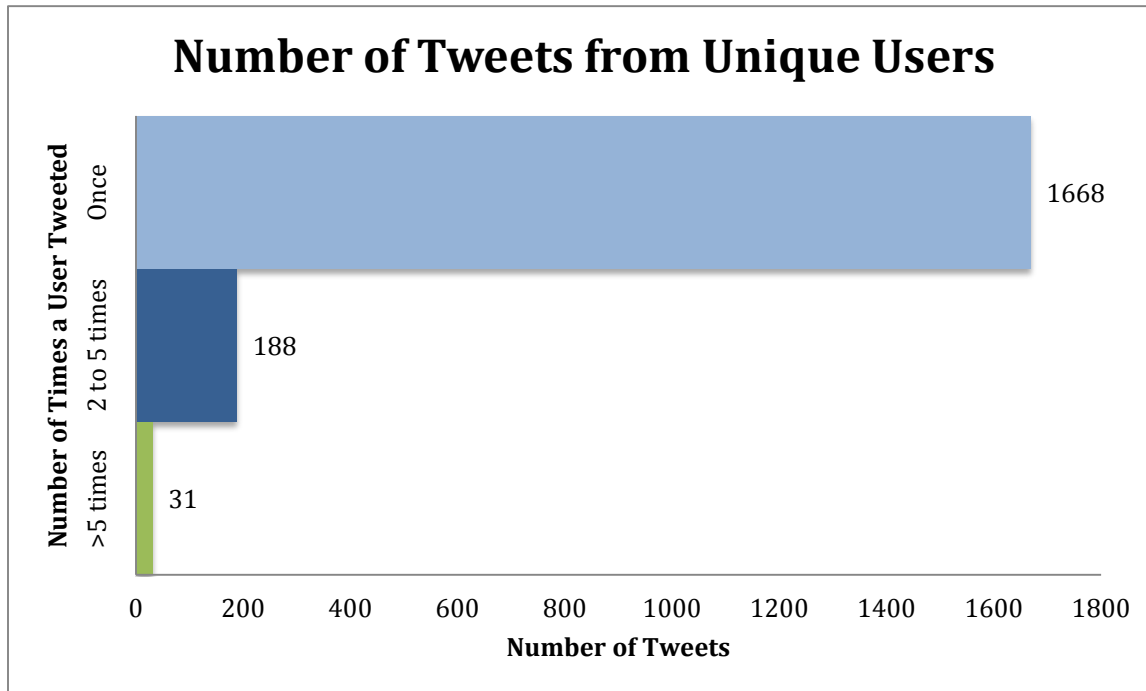
Table 1 Continued		articles about HPV or Gardasil should be included in this category.  0 – No 1 – Yes	vaccines do not lead teen girls to risky sex. Via @USATODAY <a href="http://t.co/qguNIPGV1a">http://t.co/qguNIPGV1a</a>
-------------------	--	--	---

The final codebook that was established by the two coders after a desirable inter-coder agreement was reached – a total of nine binary variables are listed. The description provides a definition of what coders were to identify, what the two binary categorizations stood for, keywords to use as a guide, and any other specifications to be mindful of. The examples listed are taken directly from the sample of tweets coders used from the dataset.

**Table 2 – Prevalence of Variables**

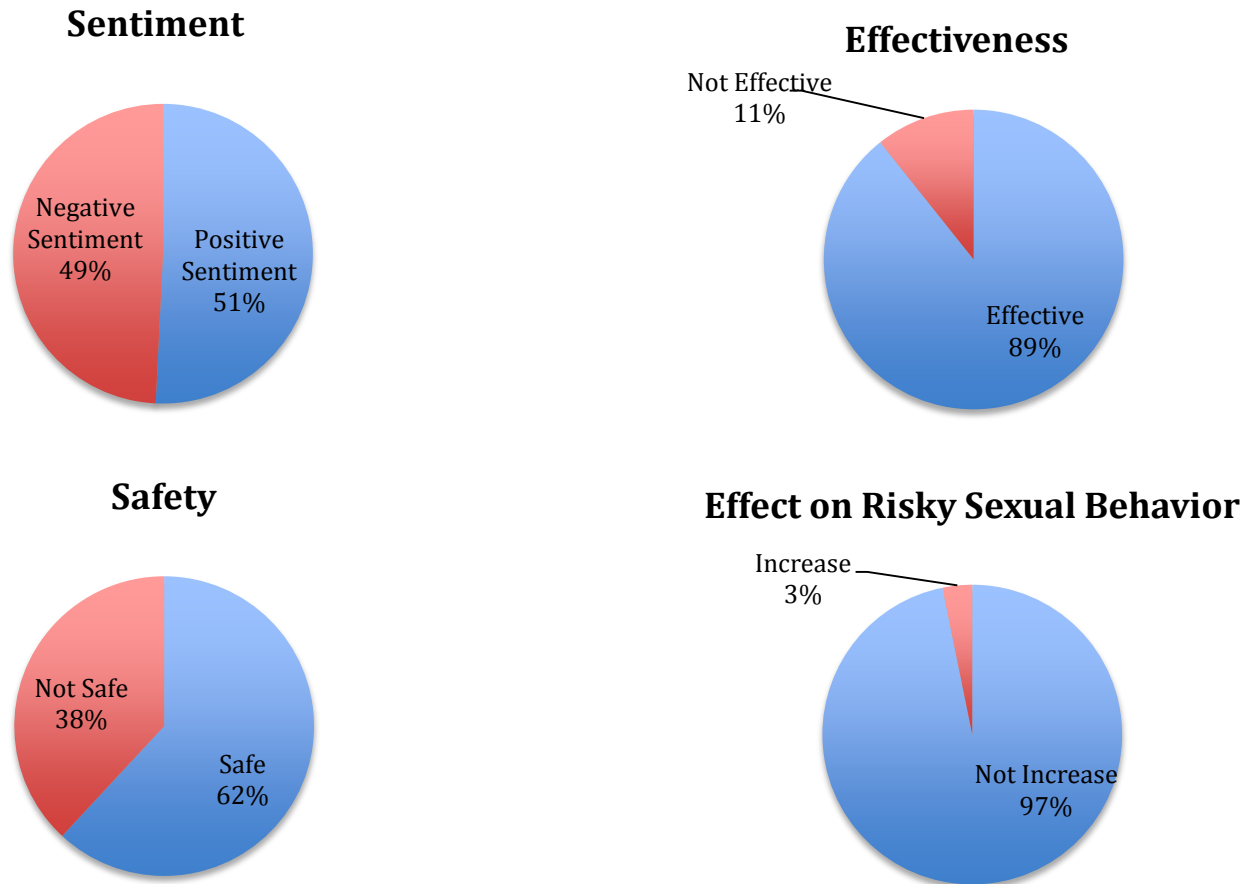
Variable	Prevalence	Example (Re-tweets/Favorites)
Positive Sentiment	5.19%	"STIs and risky behaviors don't increase after HPV Vaccination. AKA- no reason to not vaccinate!" (0/0)
Negative Sentiment	5.03%	"Gardasil: The Decision We Will Always Regret #CDCwhistleblower (0/0)
Safety	11.13%	"Just how safe is HPV vaccine? After ~700,000 doses in Ontario, looks pretty safe. #Gardasil" (38/21)
Effectiveness	5.94%	"@CDCSTD: #HPV vax helps prevent HPV types that cause most cervical #cancer #themoreyouknow" (2/0)
Sexual Behavior	28.25%	"HPV vaccine does not increase rates of STIs in adolescent females." (0/0)
News/Media	41.71%	"@TorontoStar Publisher John Cruikshank apologizes for misleading #Gardasil coverage. 'We failed. We let people down.'" (110/42)
Legal and Policy	1.43%	"Is it violating liberty for a state to offer free HPV vaccines at school, with parental opt in?" (0/0)
Barriers to Vaccination	0.90%	"Barriers to HPV vaccination: MedlinePLUS Health News Video" (1/1)
9-valent Vaccine	0.58%	"One reason Merck could possibly have for DOUBLING the amount of ALUMINUM in the new Gardasil 9 shot is to kill faster." (5/2)
Parental Attitude	0.32%	"HOW COME NOW..... THEY RECOMMEND THE HPV VACCINE FOR BOYS?? Last time I checked MY SON HAD NO OVARIES! #CDCwhistleblower #vaccines" (0/0)

This table is a representation of how prevalent each codebook variable was in our sample of 1,887 relevant tweets. The percentage listed represents the proportion of the relevant tweets that were categorized under the listed variable. For example, 98 of the 1,887 relevant tweets were marked at 'positive', which represents 5.19% of the relevant sample. Sample tweets that fell in to each category are listed, along with the number of times such tweets were re-tweeted or marked as a favorite.



**Figure 1 – Number of Tweets from Unique Users**

**This graph represents the composition of the population our sample of tweets were collected from. Of the 1,887 relevant tweets, 1,668 (88.39%) were from unique users or those who only represented one tweet from our sample. 188, or 9.96%, of the tweets came from users who posted between two to five times, and 31, or 1.64%, came from users who tweeted more than five times.**



**Figure 2 – Binary Variables**

This figure depicts the split in opinions within our sample of a few of the different binary variables that were coded for. Of the tweets that expressed sentiment, 51% were positive and 49% were negative. Among the tweets that discussed safety, 38% labeled the vaccine unsafe, while 62% labeled it safe. Of those that debated the effectiveness of Gardasil, 11% considered the vaccine ineffective while 89% said it was effective. Lastly, of the tweets that considered the effect of the vaccine on risky sexual behavior, 3% claimed that it did increase risky sexual behavior while 97% supported that it did not.

## Bibliography

1. Ghittoni, R; Accardi, R; Chiocca, S; Tommasino, M (2015), "Role of human papillomaviruses in carcinogenesis", *Ecancermedicalsecience* 9 (526)
2. Stanley, Margaret A; Winder, David M; Sterling, Jane C; Goon, Peter KC (2012). "HPV infection, anal intra-epithelial neoplasia (AIN) and anal cancer: current issues". *BMC Cancer* 12 (1): 398.
3. "Genital HPV Infection Fact Sheet". Centers for Disease Control and Prevention (CDC). April 10, 2008.
4. "HPV — the Shy Virus" (radio program). Sound print. 6 December 2008.
5. Boshart M, Gissmann L, Ikenberg H, Kleinheinz A, Scheurlen W, zur Hausen H; Gissmann; Ikenberg; Kleinheinz; Scheurlen; Zur Hausen (May 1984). "A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer". *The EMBO Journal* 3 (5): 1151–7.
6. "The Ethics of vaccination". Springer Publishing Company.
7. Markowitz, L. E.; Dunne, E. F.; Saraiya, M.; Lawson, H. W.; Chesson, H.; Unger, E. R.; Centers for Disease Control Prevention (CDC); Advisory Committee on Immunization Practices (ACIP) (2007). "Quadrivalent Human Papillomavirus Vaccine: Recommendations of the Advisory Committee on Immunization Practices (ACIP)"(PDF). *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports / Centers for Disease Control* 56 (RR-2): 1–24.
8. "FDA approves Gardasil 9 for prevention of certain cancers caused by five additional types of HPV" (press release). 10 December 2014.
9. "Genital HPV Infection - Fact Sheet." *Centers for Disease Control and Prevention*. Centers for Disease Control and Prevention, 23 Feb. 2015. Web. 28 May 2015.
10. Barth, Karen R., Robert L. Cook, Julie S. Downs, Galen E. Switzer, and Baruch Fischhoff. "Social Stigma and Negative Consequences: Factors That Influence College Students' Decisions to Seek Testing for Sexually Transmitted Infections." *Journal of American College Health* 50.4 (2002): 153-59. Web.
11. Yeganeh, Nava, Donna Curtis, and Alice Kuo. "Factors Influencing HPV Vaccination Status in a Latino Population; and Parental Attitudes towards Vaccine Mandates." *Vaccine* 28.25 (2010): 4186-191. Web.



12. Perkins, Rebecca B., Natalie Pierre-Joseph, Cecilia Marquez, Sandra Iloka, and Jack A. Clark. "Why Do Low-Income Minority Parents Choose Human Papillomavirus Vaccination for Their Daughters?" *The Journal of Pediatrics* 157.4 (2010): 617-22. Web.
13. Chando, Shingisai, Jasmin A. Tiro, T. Robert Harris, Sarah Kobrin, and Nancy Breen. "Effects of Socioeconomic Status and Health Care Access on Low Levels of Human Papillomavirus Vaccination Among Spanish-Speaking Hispanics in California." *Am J Public Health American Journal of Public Health* 103.2 (2013): 270-72. Web.
14. Ciolli, Anthony. "Mandatory School Vaccinations: The Role of Tort Law." *The Yale Journal of Biology and Medicine* 81.3 (2008): 129–137. Print.
15. Dredze, Mark. "How Social Media Will Change Public Health." *IEEE Intelligent Systems IEEE Intell. Syst.* 27.4 (2012): 81-84. Web.
16. McNab, Christine. "What Social Media Offers to Health Professionals and Citizens." *Bulletin of the World Health Organization Bull World Health Org* 87.8 (2009): 566. *Clinical Key*. Web.
17. Vance, Karl, William Howe, and Robert P. Dellavalle. "Social Internet Sites as a Source of Public Health Information." *Dermatologic Clinics* 27.2 (2009): 133-36. Web.
18. Kaplan Andreas M., Haenlein Michael (2011). "The early bird catches the news: Nine things you should know about micro-blogging" (PDF). *Business Horizons*, 54(2).. Retrieved June 5, 2014
19. Aichner, T. and Jacob, F. (March 2015). "Measuring the Degree of Corporate Social Media Use". *International Journal of Market Research* 57 (2): 257–275. Raybaut P, Nyo G. Python(x,y). 2014. Available at: <http://code.google.com/p/pythonxy/>.
20. "Twitter MAU Were 302M For Q1, Up 18% YoY - Twitter (NYSE:TWTR) | Benzinga]. April 28, 2015.
21. Thackeray, Rosemary, Brad L. Neiger, Amanda K. Smith, and Sarah B Van Wagenen. "Adoption and Use of Social Media among Public Health Departments." *BMC Public Health* 12.1 (2012): 242. Web.
22. Merchant, Raina M., Stacy Elmer, and Nicole Lurie. "Integrating Social Media into Emergency-Preparedness Efforts." *New England Journal of Medicine N Engl J Med* 365.4 (2011): 289-91. Web.

23. Colditz, J. "World Vaping Day: Contextualizing the Phenomenon of Vape Culture in Online Social Media." (n.d.): n. pag. Web.
24. Paul, M., and M. Dredze. "You Are What You Tweet: Analyzing Twitter for Public Health." *Http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2880/3264*. John Hopkins University, 2011. Web. 11 June 2015.
25. Chew, Cynthia, and Gunther Eysenbach. "Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak." *PLoS ONE* 5.11 (2010): n. pag. Web.
26. Boehner, C.W., Howe, S.R., Bernstein, D.I., Rosenthal, S.L., 2003. Viral sexually transmitted disease vaccine acceptability among college students. *Sex. Transm. Dis.* 30, 774–778.
27. Lambert, E.C., 2001. College students' knowledge of human papillomavirus and effectiveness of a brief educational intervention. *J. Am. Board Fam. Med.* 14, 178–183.
28. Gerend, M.A., Lee, S.C., Shepherd, J.E., 2006. Predictors of human papillomavirus vaccination acceptability among underserved women. *Sex. Transm. Dis.* 33, 1–4.
29. Davis, K., Dickman, E.D., Ferris, D., Dias, J.K., 2004. Human papillomavirus vaccine acceptability among parents of 10- to 15-year-old adolescents. *J. Low. Genit. Tract Dis.* 8, 188–194.
30. Constantine, N.A., Jerman, P., 2007. Acceptance of human papillomavirus vaccination among Californian parents of daughters: a representative statewide analysis. *J. Adolesc. Health* 40, 108–115.
31. Slomovitz, B.M., Sun, C.C., Frumovitz, M., Soliman, P.T., Schmeler, K.M., Pearson, H.C., Berenson, A., Ramirez, P.T., Lu, K.H., Bodurka, D.C., 2006. Are women ready for the HPV vaccine? *Gynecol. Oncol.* 103, 151–154.
32. Zimet, G.D., Mays, R.M., Winston, Y., Kee, R., Dickes, J., Su, L., 2000. Acceptability of human papillomavirus immunization. *J. Womens Health Gend.-Based Med.* 9, 47–50.
33. Dempsey, A.F., Zimet, G.D., Davis, R.L., Koutsky, L., 2006. Factors that are associated with parental acceptance of human papillomavirus vaccines: a randomized intervention study of written information about HPV. *Pediatrics* 117, 1486–1493.
34. Twitter Inc. Public API. 2014. Available at: <http://dev.twitter.com/streaming/public>.
35. Raybaut P, Nyo G. Python(x,y). 2014. Available at: <http://code.google.com/p/pythonxy/>.
36. McGrath R. Twython. 2014. Available at: <http://pypi.python.org/pypi/twython>.

37. Morstatter F, Pfeffer J, Liu H, Carley KM. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's Firehose. doi:10.1.1.300.2214.
38. Fleiss JL. The measurement of interrater agreement. In: *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons; 1981:211–236.
39. Barbosa L, Feng J. Robust sentiment detection on Twitter from biased and noisy data. In: *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. COLING '10. Stroudsburg, PA, USA: Association for Computational Linguistics; 2010:36–44. Available at: <http://dl.acm.org/citation.cfm?id=1944566.1944571>.
40. Efron M, Winget M. Questions are content: A taxonomy of questions in a microblogging environment. In: *Proceedings of the ASIST Annual Meeting*. Vol 47.; 2010. doi:10.1002/meet.14504701208.
41. Bifet A, Frank E. Sentiment Knowledge Discovery in Twitter Streaming Data. In: Pfahringer B, Holmes G, Hoffmann A, eds. *Discovery Science SE - 1*. Vol 6332. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2010:1–15. doi:10.1007/978-3-642-16184-1\_1.
42. Zhao WX, Jiang J, Weng J, et al. Comparing Twitter and Traditional Media Using Topic Models. In: *Proceedings of the 33rd European Conference on Advances in Information Retrieval*. ECIR'11. Berlin, Heidelberg: Springer-Verlag; 2011:338–349. Available at: <http://dl.acm.org/citation.cfm?id=1996889.1996934>.
43. Miller, Jake. "HMS." *HPV Vaccination Not Linked to Riskier Sex*. Harvard Medical School, 9 Feb. 2015. Web. 29 June 2015.
44. "A Note from the Publisher | Toronto Star." *Thestar.com*. The Toronto Star, 20 Feb. 2015. Web. 29 June 2015.
45. "Gene-Eden-VIR: Official Website." *Gene-Eden-VIR: Official Website*. N.p., n.d. Web. 29 June 2015.
46. Guichon, Juliet, and Rupert Kaul. "Science Shows HPV Vaccine Has No Dark Side | Toronto Star." *Thestar.com*. The Toronto Star, 11 Feb. 2015. Web. 29 June 2015.
47. Gunter, J. "Toronto Star Claims HPV Vaccine Unsafe. Science Says the Toronto Star Is Wrong." *Dr Jen Gunter*. N.p., 05 Feb. 2015. Web. 29 June 2015.