

Characterizing HPV Vaccine Discourse On Reddit

Yuki Lama¹, Dian Hu², Sandra Crouse Quinn, PhD¹, Amelia Jamison¹, David A. Broniatowski, PhD²

¹University of Maryland, College Park, MD; ²George Washington University, Washington, D.C.

Introduction:

Approximately 23,000 women and 15,793 men in the United States are affected by human papilloma virus (HPV) related cancers. Many of the infections could have been prevented through vaccination [2]. The CDC has recommended a 3-dose vaccination series as a safe and effective method of protecting against HPV strains associated with cervical and other cancers for girls since June 2006 and for males since October 2011. However, vaccination rates remain low. Strengthening uptake efforts is a preeminent national public health concern, as demonstrated by the Healthy People 2020 objective to increase HPV three-doses vaccination series completion for adolescents aged 13–15 to 80% by the year 2020. In order to promote HPV vaccination through public health communication efforts, the mechanisms in which individuals utilize resources related to health information must be identified and understood. Surveillance of social media data can provide an alternative approach to observe patterns of information dissemination and knowledge exchange as related to HPV vaccination behaviors [1]. Understanding extant online discussions on HPV vaccination is pivotal in developing concerted, tailored health communication efforts to enhance HPV vaccination rates. There is a growing number of studies contributing to the literature addressing the nexus of HPV and social media data, but to date, no published study has examined Reddit content related to HPV vaccination.

Objective:

We seek to observe the following trends using Reddit messages: (1) how the HPV vaccine is characterized on Reddit— cancer risk vs. sexual behavior concerns; (2) how these discussions change over time.

Methods:

Based on the research question, all public Reddit comments were gathered from Jan 2006 to Dec 2015 to examine temporal trends in the discourse on the HPV vaccine, using a custom scraper implemented in Python. The JSON library was used to process raw data and the NLTK library was used for the purpose of tokenizing and stemming. DateTime library was utilized to categorize messages based on their timestamp. During the extraction process, messages were considered potentially relevant for further analysis if they contained the English strings “hvp” and “vaccin”. After the filtering process, 22,750 potential HPV-vaccine related messages were identified. Qualitative analyses of a subset of the messages were used to complement and guide trained classifiers. Two annotators manually evaluated 100 messages using a qualitative codebook developed (see below). The evaluation procedure attempts to categorize whether the message is discussing (1) cancer risks and (2) sexual behavior. Basic demographic information including age and gender would be collected to discern the user base as well as any biases inherent to the platform. Hand annotation is an iterative process with the aim of assembling 100 “gold” messages that best represent the coded themes of interest mentioned. On the first round of annotation, raters coded approximately 50% of the messages related to sexual behavior (inter-rater reliability: $\kappa = 0.838$; SE 0.055) and 59-62% of the messages related to cancer risk (inter-rater reliability: $\kappa = 0.726$; SE 0.070). Upon completion of this project, manually annotated messages will be split into training data and test data; 90% of the manually annotated messages would be used as training data and 10% of the messages would be kept as test data. Consistent with the categories listed in the codebook, we will next build a classifier that can automatically catalog the messages using the codes identified above. In other words, the classifier will be able to analyze the entire sample, based on the “gold” messages that were annotated, with a 75% accuracy to ultimately examine HPV vaccine trends on Reddit over time.

Conclusion:

Using a combination of qualitative analyses and natural language processing techniques, this study investigates the characterization of HPV vaccination discourse on Reddit over time, with particular focus on cancer prevention and sexual behavior. Social media can be a tool to examine HPV discourse over time in order to inform the public health promotion goal of increasing HPV vaccination uptake. Public health communication can harness the rapid, low-cost dissemination capabilities of social media to deliver timely, accurate health information to the general public.

Table 1. Qualitative codebook and sample messages

DOMAIN:	DESCRIPTION:	EXAMPLE:
Cancer Risk	Messages are related to HPV and relationship to cancer and HPV vaccine. This includes concerns related to cancer, awareness of cancer risk. Comments can be any related type of cancer: cervical, throat, penile, anal, etc.	<p><i>it's not just genital warts, the hpv vaccine can prevent cervical cancer. amazing stuff.</i></p> <p><i>the vaccines don't prevent all causes of cervical cancer, they protect against some strains of hpv which can cause cervical cancer. also, they only started giving that vaccine to women over 25 recently.</i></p>
General Sexual Behavior	Messages are related to HPV and sexual activity including discussions of age of sexual debut, risky sexual behavior, multiple sex partners, associations with promiscuity, etc. Messages need an explicit connection to HPV. Can include screening behavior if it is related to sexual activity.	<p><i>seeing as he has already claimed that the hpv vaccine would make his daughters promiscuous and tried to stop it being made compulsory for girls, it would not surprise me at all.</i></p> <p><i>thanks for clearing that up. i'm a guy, so it's a little tough to wrap my head around. i might get the vaccine just as a preventative measure. i've only had one sexual partner in my history and we were both virgins, so i'm certain i'm clear. not that hpv is a death sentence or anything, but i'd rather not have to deal with it, and so i don't unknowingly spread it to others in the case that i did get it.</i></p>

References:

1. Dunn, A. G., Surian, D., Leask, J., Dey, A., Mandl, K. D., & Coiera, E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. *Vaccine*. 2017; 35(23), 3033–3040. <https://doi.org/10.1016/j.vaccine.2017.04.060>
2. Reagan-Steiner S, Yankey D, Jeyarajah J, et al. National, Regional, State, and Selected Local Area Vaccination Coverage Among Adolescents Aged 13–17 Years — United States. *MMWR Morb Mortal Wkly Rep* 2016; 65:850–858. DOI: <http://dx.doi.org/10.15585/mmwr.mm6533a4>