

HPV Vaccine Conversations on Reddit (2016-2025): Shifts in Public Sentiment and Themes in the COVID-19 Era

Marcin Marciniak, Sean A. Setzen, Deeya Bhattacharya, Jamie Masliah, Robin Powszok, Erich M. Sturgis, Vanessa C. Stubbs, Mihir K. Bhayani

Submitted to: Journal of Medical Internet Research
on: September 05, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	27
Figures	28
Figure 1.....	29
Figure 2.....	30
Multimedia Appendixes	31
Multimedia Appendix 1.....	32
Multimedia Appendix 2.....	32
Multimedia Appendix 3.....	32
Related publication(s) - for reviewers eyes onlies	33
Related publication(s) - for reviewers eyes only 0.....	33
Related publication(s) - for reviewers eyes only 0.....	33

HPV Vaccine Conversations on Reddit (2016–2025): Shifts in Public Sentiment and Themes in the COVID-19 Era

Marcin Marciniak^{1, 2} BS; Sean A. Setzen^{1, 3} MD; Deeya Bhattacharya⁴ BS; Jamie Masliah¹ MD; Robin Powszok¹ MD; Erich M. Sturgis⁵ MPH, MD; Vanessa C. Stubbs¹ MD; Mihir K. Bhayani¹ MD

¹ Department of Otorhinolaryngology-Head and Neck Surgery Rush University Medical Center Chicago US

² Rosalind Franklin University of Science and Medicine North Chicago US

³ Department of Otorhinolaryngology-Head and Neck Surgery St. Luke's University Health Network Bethlehem US

⁴ Rush University Medical College Chicago US

⁵ Department of Otorhinolaryngology-Head and Neck Surgery Baylor College of Medicine Houston US

Corresponding Author:

Mihir K. Bhayani MD

Department of Otorhinolaryngology-Head and Neck Surgery
Rush University Medical Center
1611 W Harrison St Suite 550
Chicago
US

Abstract

Background: The HPV vaccine is a key preventative tool against HPV-related cancers, yet vaccination rate remains at 66% in the U.S. As social media becomes an increasingly popular avenue for health discussions, platforms like Reddit offer unique opportunities to understand public sentiment and discourse surrounding vaccination. This study analyzed HPV vaccine-related discussions on Reddit to examine evolving themes and sentiment, particularly in the context of the COVID-19 pandemic.

Objective: This study aimed to analyze HPV vaccine-related discussions on Reddit to (1) identify dominant discussion themes, (2) assess overall sentiment toward the HPV vaccine, and (3) examine changes in themes and sentiment before and after the start of the COVID-19 pandemic.

Methods: A Python-based script was used to scrape Reddit posts and comments referencing the HPV vaccine from multiple subreddits between 2016 and 2025. Posts were manually labeled as either intentional or incidental, with intentional entries further categorized into ten thematic groups. Sentiment analysis was conducted using the VADER algorithm. Temporal trends were assessed using pre- and post-COVID-19 stratification. Chi-square tests, Mann-Whitney U tests, and linear regression were used for statistical analysis.

Results: Of 4,235 collected entries, 2,801 were included in the final analysis. The most common themes were factual claims (25.1%), personal experiences (20.4%), and offering advice (16.6%). Sentiment shifted significantly post-COVID, with positive sentiment increasing from 44.6% to 52.4% ($P=.01$). Themes related to vaccine hesitancy and cultural influences shifted from negative to neutral sentiment, though not statistically significant. Median time from post to first comment was 136 minutes, suggesting high engagement.

Conclusions: HPV vaccine discourse on Reddit has become more positive and experience-driven over time. These insights underscore Reddit's role in shaping health perceptions and offer opportunities for targeted public health engagement in the digital age.

(JMIR Preprints 05/09/2025:83558)

DOI: <https://doi.org/10.2196/preprints.83558>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [JMIR Publications](#)

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in [JMIR Publications](#)

Original Manuscript

Original Paper

HPV Vaccine Conversations on Reddit (2016–2025): Shifts in Public Sentiment and Themes in the COVID-19 Era

Marcin Marciniak BS^{1,2}, Sean A. Setzen MD^{1,3}, Deeya Bhattacharya, BS⁴, Jamie Masliah MD¹, Robin Powszok MD¹, Erich M. Sturgis MD, MPH⁵, Vanessa C. Stubbs MD¹, Mihir K. Bhayani MD¹

- 1.) Department of Otorhinolaryngology-Head and Neck Surgery, Rush University Medical Center, Chicago, IL, USA.
- 2.) Chicago Medical School at Rosalind Franklin University of Medicine and Science, North Chicago, Illinois, U.S.A.
- 3.) Department of Otorhinolaryngology-Head and Neck Surgery, St. Luke's University Health Network, Bethlehem, Pennsylvania, U.S.A.
- 4.) Rush University Medical College, Chicago, IL, USA.
- 5.) Department of Otorhinolaryngology-Head and Neck Surgery, Baylor College of Medicine, Houston, TX, USA.

Conflicts of Interest/Disclosures: The authors declare no conflicts of interest or disclosures.

Funding: No funding was obtained for the completion of this project.

Corresponding Author:

Mihir K. Bhayani, MD, FACS

Director of the Rush Head and Neck Cancer Research Program and Salivary Gland Program,
Associate Professor

Department of Otorhinolaryngology-Head and Neck Surgery

Rush University Medical Center

1611 W Harrison St., Suite 550

Chicago, IL 60612

Email: mihir_bhayani@rush.edu

Phone Number: (312) 942-6100

Fax Number: (312) 942-6225

Keywords: Papillomavirus Vaccines; Papillomavirus Infections; Social Media; Sentiment Analysis; COVID-19

Abstract

Background

The HPV vaccine is a key preventative tool against HPV-related cancers, yet vaccination rate remains at 66% in the U.S. As social media becomes an increasingly popular avenue for health discussions, platforms like Reddit offer unique opportunities to understand public sentiment and discourse surrounding vaccination. This study analyzed HPV vaccine-related discussions on Reddit to examine evolving themes and sentiment, particularly in the context of the COVID-19 pandemic.

Objective

This study aimed to analyze HPV vaccine-related discussions on Reddit to (1) identify dominant discussion themes, (2) assess overall sentiment toward the HPV vaccine, and (3) examine changes in themes and sentiment before and after the start of the COVID-19 pandemic.

Methods

A Python-based script was used to scrape Reddit posts and comments referencing the HPV vaccine from multiple subreddits between 2016 and 2025. Posts were manually labeled as either intentional or incidental, with intentional entries further categorized into ten thematic groups. Sentiment analysis was conducted using the VADER algorithm. Temporal trends were assessed using pre- and post-COVID-19 stratification. Chi-square tests, Mann-Whitney U tests, and linear regression were used for statistical analysis.

Results

Of 4,235 collected entries, 2,801 were included in the final analysis. The most common themes were factual claims (25.1%), personal experiences (20.4%), and offering advice (16.6%). Sentiment shifted significantly post-COVID, with positive sentiment increasing from 44.6% to 52.4% ($P=.01$). Themes related to vaccine hesitancy and cultural influences shifted from negative to neutral sentiment, though not statistically significant. Median time from post to first comment was 136 minutes, suggesting high engagement.

Conclusion

HPV vaccine discourse on Reddit has become more positive and experience-driven over time. These insights underscore Reddit's role in shaping health perceptions and offer opportunities for targeted public health engagement in the digital age.

Introduction

Human papillomavirus (HPV) is a sexually transmitted virus and a well-established risk factor for various cancers in both men and women.¹ Most commonly associated with cervical cancer, HPV has also been increasingly linked to anogenital and oropharyngeal cancers.¹ High-risk HPV strains, particularly types 16 and 18, are responsible for over 70% of these cases, while low-risk strains cause benign conditions such as papillomas and warts.^{2,3}

Vaccines have been developed to protect against both high-risk and low-risk HPV strains, reducing the likelihood of persistent infection. The first HPV vaccine, Gardasil, was approved by the U.S. Food and Drug Administration (FDA) in 2006 and provided protection against HPV 6, 11, 16, 18.⁴ Subsequent vaccines, such as Gardasil 9, expanded coverage to include five additional high-risk strains (HPV 31, 33, 45, 52, and 58), offering broader protection against HPV-associated cancers.⁵ Despite these advancements, vaccination coverage remains suboptimal, with only about 66% of

eligible individuals completing the full vaccine series by age 17, and approximately 78% receive at least one dose.⁶ Given the largely preventable nature of HPV-associated cancers, global initiatives including the World Health Organization (WHO) Cervical Cancer Elimination Initiative, have emphasized increasing HPV vaccination as a key public health priority.⁷

Unlike vaccines routinely required for school attendance, such as those for Tdap, Hepatitis B, and meningococcal disease, the HPV vaccine encounters distinct challenges. Mandates for HPV vaccination exist in only a few states, and even these often include opt-out provisions that hinder widespread uptake.⁸ Since its introduction, the vaccine has also been embroiled in controversy, with public hesitancy fueled by concerns regarding its safety, efficacy, and perceived association with sexual activity.^{8,9} The COVID-19 pandemic may have exacerbated this skepticism, as heightened public discourse around vaccines amplified misinformation, emotional responses, and confusion.¹⁰

While healthcare providers typically rely on evidence-based guidelines for vaccine advocacy, this approach can sometimes overlook the personal experiences, specific concerns, and misconceptions held by patients. Addressing these individual perspectives is crucial both for improving HPV vaccination rates and for bridging the divide between scientific recommendations and public apprehension. In this context, social media platforms provide rich readily available datasets for exploring patient viewpoints, offering advantages over traditional survey methods which can be resource-intensive and unable to capture spontaneous, real-time discussions.¹¹

Although previous research has extensively analyzed HPV vaccine discussions on Twitter—a platform popular among adolescents and young adults, who are key demographics for the vaccine—less attention has been paid to corresponding discussions on Reddit.^{9,11–18} Earlier work by Lama et al. examined Reddit discourse from 2007 to 2015 using topic modeling, identifying themes like political debate, sexual behavior, and vaccination schedules.¹⁹ More recent work by Du et al. and Meci et al. has explored misinformation prevalence and conducted sentiment analysis of Reddit posts, respectively, but both studies were limited by narrow thematic scopes and did not assess how discourse evolved in response to the COVID-19 pandemic.^{20,21}

Consequently, there is a gap in understanding the longitudinal evolution of HPV vaccine discourse on Reddit, particularly concerning shifts in themes and sentiment surrounding major public health events like the COVID-19 pandemic, analyzed through detailed manual coding. This study addresses this gap by analyzing both Reddit posts and comments from 2016 to 2025 using a granular, ten-category manual coding scheme. It uniquely examines how discussion themes and overall sentiment changed before and after the onset of the COVID-19 pandemic.

Reddit offers a distinct environment for such analysis. Unlike platforms often linked to identifiable user accounts, Reddit facilitates anonymous, topic-centered conversations, creating an open forum for exploring public sentiment candidly. Dedicated communities, such as the r/HPV subreddit with over 22,000 members, function as valuable spaces where individuals share personal stories, seek advice, and voice concerns about the HPV vaccine. Analyzing these interactions can illuminate prevalent misconceptions, identify barriers to vaccination, and reveal opportunities for more effective public health communication.

Therefore, this study aims to characterize the discourse surrounding the HPV vaccine on Reddit by identifying recurring discussion themes and associated sentiment, thereby exploring potential factors contributing to suboptimal uptake. Furthermore, it investigates the evolution of these discussions, especially influenced by the COVID-19 pandemic, to offer insights that can inform more effective and empathetic communication strategies by healthcare professionals and public health officials.

Methods

Data Collection

Posts and comments referencing the HPV vaccine were collected from multiple subreddits, along with their respective metadata. Subreddits analyzed are outlined in Table 1. Using Visual Studio Code (Version 1.90.0 (Universal)), an Integrated Development Environment developed by Microsoft, a data scraping script was written in Python (version 3.9.16) using the praw library to interact with the Reddit API and collect entries mentioning the HPV vaccine from selected subreddits (Table 1).

Table 1. Data Sources: Subreddit Communities Included and Applied Keyword Filters.

Subreddit Category	Subreddits	Keywords Applied
HPV-Focused	r/HPV, r/PreCervicalCancer, r/Warts, r/CervicalCancer	Vaccine-Related Keywords: - vaccine, vaccines, vaccinated, vaccination, vaccinations, vax, vaxxed, unvaxxed - Gardasil, cervarix - anti-vax, antivax, antivaccine, anti-vaccine, pro-vax - jab, shot, immunization
Vaccine-Focused	r/VACCINES, r/vaxxhappened, r/Vaccine, r/antivax, r/DebateVaccines, r/unvaccinated, r/skeptic, r/StopMandatoryVaccine	HPV-Related Keywords: - HPV, HPV vaccine, HPV shot - gardasil, cervarix - human papillomavirus
Medical & Science	r/AskDocs, r/medicine, r/PeterAttia, r/HeadandNeckCancer, r/Health, r/ScienceBasedParenting, r/HNSCC	HPV + Vaccine-Related Keywords: - HPV, HPV vaccine, HPV shot, gardasil, cervarix, human papillomavirus - vaccine, vaccines, vaccinated, vaccination, jab, shot, immunization - anti-vax, antivax, pro-vax
Women's Health	r/TwoXChromosomes, r/TheGirlSurvivalGuide, r/WomenOver40, r/AskWomenOver30	
General & Social	r/MensRights, r/AskGaybrosOver30, r/AskConservatives, r/askgaybros	

For each post and comment, metadata—including the subreddit name, text body, flair, date of submission, number of upvotes, post or comment ID, parent comment ID, entry type (post or comment), and anonymized author ID—was collected. Posts and comments were included if they contained at least one relevant keyword, as described below. Comments were only retained if they belonged to posts that met keyword criteria, ensuring contextual relevance to the HPV vaccine.

Keyword Mapping and Filtering

A systematic keyword mapping approach was used to identify relevant discussions. Keywords were divided into primary categories: general vaccine-related terms ("vaccine," "vaccination," "vax," "jab") and HPV-specific terms ("HPV vaccine," "Gardasil," "Cervarix") (Table 1). These keyword groups were mapped to subreddits based on thematic relevance. Subreddits specifically dedicated to HPV (e.g., r/HPV) were filtered using the general vaccine-related keyword group, whereas vaccine-

focused subreddits (e.g., r/VACCINES, r/vaxxhappened) were filtered using the HPV-specific keyword group. Subreddits with broader medical or health focuses (e.g., r/AskDocs) used both keyword groups to capture relevant content.

To improve filtering accuracy, text was converted to lowercase for case-insensitive matching. Keywords were processed to prioritize longer, more specific terms (e.g., "HPV vaccine" before "HPV") to prevent mismatches where shorter keywords might appear within unrelated longer words or phrases.

Data Labeling and Categorization

Retrieved Reddit entries were imported into a Microsoft Excel spreadsheet, organizing metadata into corresponding columns. Initially, each entry was classified into one of two groups: incidental mentions or intentional discussions. Incidental mentions were those containing a relevant keyword but not primarily focused on the HPV vaccine (e.g., brief mentions within a broader vaccination history discussion). In contrast, intentional discussions explicitly addressed the HPV vaccine as a central topic.

Intentional entries underwent further classification into ten thematic subcategories: offering advice/recommendations, seeking advice/recommendations, general questions, vaccine hesitancy, social/cultural influences, personal experiences, resource or data sharing, barriers to accessibility, factual claims, and other. The stepwise data labeling process is illustrated in Figure 1. Specifically, advice themes captured entries either offering ("You should get the HPV vaccine") or seeking ("Should I get the HPV vaccine?") recommendations. General questions involved inquiries about specific vaccine details, like efficacy or dosing schedules. Vaccine hesitancy included expressions of concern about safety or side effects, as well as content perpetuating misinformation or myths. Social/cultural influences covered discussions on how factors like family, religion, or social stigma impacted vaccine attitudes. Personal experiences detailed users' firsthand accounts related to vaccination, side effects, or post-vaccination HPV diagnoses. Resource or data sharing comprised entries providing links, research articles, or other information. Barriers to accessibility described difficulties obtaining the vaccine due to cost, insurance, or availability issues. Factual claims presented information as fact (e.g., "Gardasil protects against HPV types 6, 11, 16, and 18"), irrespective of accuracy, including potentially false claims presented informatively. Entries not fitting these definitions were categorized as other. Detailed definitions and examples for each theme are provided in Table 2.

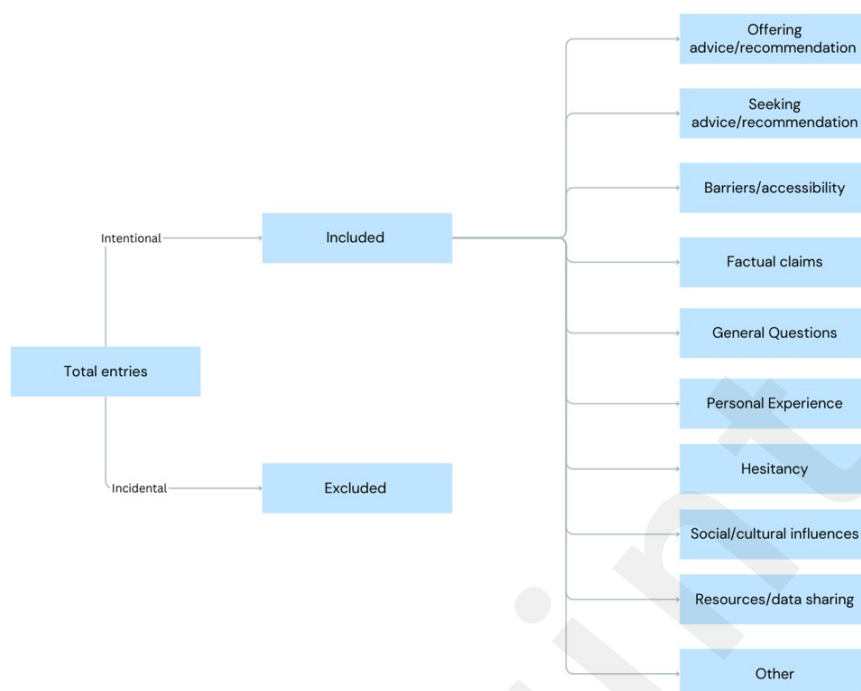


Figure 1. Flowchart of Post and Comment Screening and Thematic Coding. Reddit entries were first screened for relevance and classified as Intentional (explicitly addressing HPV vaccination) or Incidental (unrelated content, excluded). Intentional posts were then subjected to thematic coding into ten mutually exclusive categories.

Table 2. Thematic Codebook: Thematic Categories, Definitions, and Sample Excerpts.

Theme	Definition	Examples
Advice/Recommendations (Offering)	Posts that explicitly offer advice or recommendations regarding HPV vaccination to others.	“You should get vaccinated”, “My doctor recommended getting the vaccine”
Advice/Recommendations (Seeking)	Posts that request advice or opinions from others about receiving the HPV vaccine.	“Should I get the vaccine?”, “Should I encourage my partner to get the vaccine?”
Barriers/Accessibility	Posts that describe challenges in accessing the HPV vaccine, including financial, insurance, geographic, or age-related limitations.	“I wanted the vaccine 3 months before I was 26, insurance said no, my age disqualified coverage, but I hadn’t turned 26 yet”, “My daughter got the shot after she moved out and the first shot was over 400 dollars. I told her about in network and the 2nd was free.”
Factual Claims	Statements presented as facts, regardless of accuracy, typically lacking overt opinion or emotion. Intended to convey information or assert knowledge.	“Gardasil 9 protects against nine types of human papillomavirus (HPV): 6, 11, 16, 18, 31, 33, 45, 52, and 58”, “There’s zero data on actual cervical cancer rates declining due to HPV vaccines.”
General Questions	Posts that inquire about specific aspects of the HPV vaccine, including dosing, efficacy, safety, or eligibility.	“Is two shots enough to protect you?”, “Why the difference between US CDC 3-dose and UK NHS 2-dose recommendation for HPV vaccination for older adults up to age 45?”
Hesitancy	Posts expressing concerns, skepticism, or fear about HPV vaccine safety, efficacy, or perceived risks. May include conspiracy or mistrust themes.	“I know someone who periodically needs to have her skull drained of excess fluid because of the HPV vaccine. They almost died after taking it due to fluid build up around her brains. I stupidly took it as an adult at 36 and have had autoimmune issues ever since.”, “And my generation, 90% of us have HPV and don’t know it and we’re fine... U think doctors don’t lie? Doctors aren’t corrupt or capable of being

		corrupted? Just use common sense. Less vaccines the better IMHO!”
Personal Experience	Narratives sharing personal or family experiences with HPV vaccination, including positive, negative, or neutral outcomes.	“I got my 1st dose of Gardasil vaccine on Thursday. I have a fever of 103.3 and I feel really sick. How are you doing?”, “I was diagnosed with HPV after almost 20 years of marriage... I’m now getting vaccinated at 43, even though I had a hysterectomy because I want to make sure I’m protected from any possible further in the future.”, “I had the vaccine aged 12, and got diagnosed with HPV at my first smear, ended up having a colposcopy to remove dodgy cells. It is not a guaranteed prevention.”
Resources/Data Sharing	Posts that share articles, external links, or statistical information to support a claim, argument, or question.	“The CDC recommends that women get the HPV vaccine through age 26, if they haven’t already been vaccinated. https://www.cancer.gov/types/cervical/causes-risk-prevention ”, “The risk of cancer for low risk strains is quite low. The high risk strains themselves comprise the majority of HPV-derived cancer. See [this table](https://www.cdc.gov/cancer/hpv/statistics/cases.htm) for a breakdown of risk of cancer type by HPV strain.”
Social/Cultural Influences	Posts discussing how culture, religion, family, or societal norms influence attitudes or decisions regarding HPV vaccination.	“I thought, because of what the Ministry of Health informed back then in my country, the HPV vaccine only worked in young girls.”, “I never got the vaccine when I was younger (due to being too young to make my own decision and my mom being against it). I’m obviously very upset but just happy to see this post and hopefully my HPV clears up and goes dormant once this is all over with”
Other	Posts that don’t fit into other categories due to ambiguity, sarcasm, or lack of context.	“At what age did you get the vaccine?”, “I’m not her, so no - her witness testimony is evidence of her claims & she’s more credible than Gardasil.”

To ensure reliability, two authors (M.M., D.B.) independently performed this manual data labeling, resolving any discrepancies through discussion and consensus.

Sentiment Analysis and Temporal Stratification

Sentiment analysis was conducted using the VADER (Valence Aware Dictionary and Sentiment Reasoner) algorithm, a lexicon and rule-based sentiment analysis tool specifically attuned to sentiments expressed in social media. VADER assigns each entry a compound score from -1 (most negative) to +1 (most positive). Based on established thresholds, entries were classified as positive (score > 0.05), neutral (score between -0.05 and 0.05), or negative (score < -0.05).

Sentiment scores were analyzed across the entire data set and within each thematic category. Due to the non-normal distribution of these scores, the overall sentiment trend was summarized using the median and interquartile range (IQR). However, mean sentiment scores were reported for individual themes to facilitate interpretation and cross-theme comparisons.

To assess the potential impact of the COVID-19 pandemic, entries were temporally stratified based on the WHO's pandemic declaration date (March 11, 2020). Entries posted before this date were designated pre-COVID, while those posted on or after were designated post-COVID.

Statistical Analysis

Descriptive statistics, including frequency counts and percentages, were used to characterize the dataset and thematic distribution. A chi-square test of independence compared theme frequencies between the pre- and post-COVID periods to identify significant changes in discussion topics over

time. Standardized residuals ($> |2|$) highlighted specific themes that were significantly over- or underrepresented in each period. To examine trends in HPV vaccine-related Reddit activity over time, a Poisson regression was conducted using annual post and comment counts from 2016 to 2025. Additional models stratified by pre- and post-COVID periods were used to compare trends before and after the pandemic declaration. Incidence rate ratios (IRRs), confidence intervals, and p-values were reported to evaluate the magnitude and significance of year-over-year changes in post volume.

Given the non-parametric nature of the sentiment scores, a Mann-Whitney U test was employed to compare overall median sentiment between the pre- and post-COVID periods. Median sentiment values and interquartile ranges were reported for each time period. Additional Mann-Whitney U tests were performed for each of the ten themes to evaluate whether sentiment scores changed significantly within individual thematic categories before and after the COVID-19 pandemic. Sentiment distributions and median values were compared for each theme pre- and post-COVID. Due to the non-parametric nature of the sentiment data and variability in group sizes, the Mann-Whitney U test was selected over parametric alternatives. Statistical significance was set at $P < .05$, but p-values were interpreted cautiously given multiple comparisons and small sample sizes in some categories.

Furthermore, a linear regression model assessed whether thematic category significantly predicted sentiment score, controlling for other themes. Using "Factual Claims" as the reference category (due to its prevalence and relatively neutral informational nature), each theme was treated as a categorical predictor. This approach allowed for comparing the sentiment associated with each theme relative to this baseline. Model fit was evaluated using R^2 and overall F-statistics.

Finally, user engagement was assessed by calculating the median response time (time elapsed from the submission of an original post to the first collected comment) across the included subreddits. This metric provided insight into the platform's interaction speed and responsiveness regarding vaccine-related inquiries. All statistical analyses were conducted in RStudio (version 2024.12.1+563), and statistical significance was defined as $P < .05$.

Ethical Considerations

Data collection adhered to Reddit's API terms of use; all retrieved data was anonymized and derived from publicly available content. The study protocol was reviewed by the Institutional Review Board of Rush University Medical Center and granted exemption.

Results

Overview

A total of 4,235 Reddit entries referencing the HPV vaccine were collected between September 13, 2016, and February 13, 2025. Of those, 2,803 were classified as intentional posts or comments. After excluding two entries from 2011 that fell outside the defined study period, 2,801 intentional entries were included in the final analysis, including 614 posts and 2,187 comments.

Factual claims were most common (703/2,801; 25.1 %), followed by personal experiences (573; 20.4 %) and advice-offering (466; 16.6 %) (Table 3). Overall, 1,447 entries (51.7%) exhibited positive sentiment, 1,052 (37.6%) negative, and 302 (10.8%) neutral. The median sentiment score across all entries was 0.13 [IQR -0.49 to 0.71]. At the theme level, advice-seeking had the highest mean sentiment (0.17 ± 0.62), whereas "Other" entries were slightly negative (-0.04 ± 0.65).

Table 3. Theme Frequencies and Mean Sentiment Scores Before versus After World Health Organization's COVID-19 Pandemic Declaration (March 11, 2020).

Theme	Pre-COVID Count (%)	Pre-COVID Sentiment (mean $\pm$ SD)	Post-COVID Count (%)	Post-COVID Sentiment (mean $\pm$ SD)	Total Count (%)	Total Sentiment (mean $\pm$ SD)
Advice/ Recommendations (Offering)	34 (12.2%)	0.05 $\pm$ 0.69	432 (17.1%)	0.12 $\pm$ 0.63	466 (16.6%)	0.12 $\pm$ 0.64
Advice/ Recommendations (Seeking)	5 (1.8%)	-0.09 $\pm$ 0.30	148 (5.9%)	0.18 $\pm$ 0.63	153 (5.5%)	0.17 $\pm$ 0.62
Barriers/ Accessibility	52 (18.7%)	0.03 $\pm$ 0.71	168 (6.7%)	0.12 $\pm$ 0.63	220 (7.8%)	0.10 $\pm$ 0.65
Factual Claims	71 (25.5%)	0.00 $\pm$ 0.62	632 (25.0%)	0.09 $\pm$ 0.65	703 (25.1%)	0.08 $\pm$ 0.65
General Questions	16 (5.8%)	-0.01 $\pm$ 0.56	199 (7.9%)	0.08 $\pm$ 0.69	215 (7.7%)	0.07 $\pm$ 0.68
Hesitancy	9 (3.2%)	-0.27 $\pm$ 0.52	132 (5.2%)	0.09 $\pm$ 0.65	141 (5.0%)	0.07 $\pm$ 0.64
Personal Experience	37 (13.3%)	0.08 $\pm$ 0.62	536 (21.2%)	0.14 $\pm$ 0.64	573 (20.4%)	0.13 $\pm$ 0.64
Resources/Data Sharing	21 (7.6%)	0.05 $\pm$ 0.70	161 (6.4%)	0.10 $\pm$ 0.66	182 (6.5%)	0.09 $\pm$ 0.67
Social/Cultural Influences	28 (10.1%)	-0.17 $\pm$ 0.73	62 (2.5%)	0.14 $\pm$ 0.68	90 (3.2%)	0.04 $\pm$ 0.71
Other	5 (1.8%)	0.11 $\pm$ 0.67	53 (2.1%)	-0.05 $\pm$ 0.65	58 (2.1%)	-0.04 $\pm$ 0.65
Total	278	0.00 $\pm$ 0.65	2,523	0.11 $\pm$ 0.65	2,801	0.10 $\pm$ 0.65

Note. Counts and percentages represent the total number of entries per theme in each period. Mean sentiment scores ($\pm$ SD) are shown alongside their respective theme frequencies.

Pre- vs. Post-COVID Volume, Themes, and Sentiment

Temporal analysis revealed increasing engagement with HPV vaccine topics over time (Figure 2). Annual post counts rose from 12 in 2016 to 132 in 2018, then accelerated after 2021, peaking at 1,337 in 2024 (Figure 2). Early 2025 data comprised 563 entries. A Poisson regression model assessing annual post and comment volume from 2016 to 2025 confirmed a significant upward trend. Each additional year was associated with a log-linear increase in count ($\beta = .41$, SE = 0.01; IRR = 1.50, 95% CI [1.48-1.52], $P < .001$), corresponding to a 50% rise in activity per year. When stratified, the pre-COVID period showed a 13% annual increase ($\beta = .12$, SE = 0.04; IRR = 1.13, 95% CI [1.04-1.23], $P = .01$), while the post-COVID period demonstrated a steeper 60% annual increase ($\beta = .47$, SE = 0.01; IRR = 1.60, 95% CI [1.56-1.65], $P < .001$).

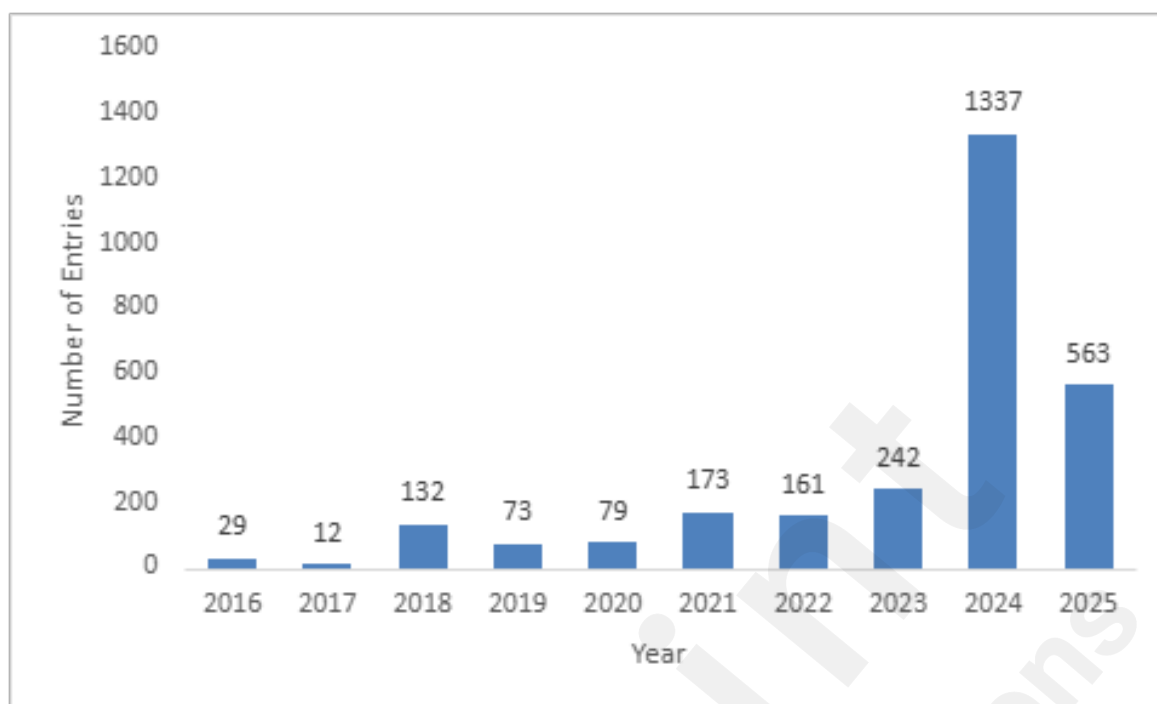


Figure 2. Annual Volume of Reddit Posts Referencing the HPV Vaccine, September 2016–February 2025. Bars indicate the total number of Reddit entries containing HPV vaccine-related keywords in each calendar year.

Of the 2,801 entries, 278 occurred pre-COVID (Sep 2016–Mar 2020) and 2,523 post-COVID (Apr 2020–Feb 2025). Theme distributions differed significantly pre- vs. post-COVID (χ^2 (df = 9, N = 2,801) = 114.47, $P < .001$). Standardized residuals showed pre-COVID overrepresentation in Barriers/Accessibility (SR = 7.09) and Social/Cultural Influences (SR = 6.83), and post-COVID overrepresentation in Personal Experience (SR = 3.11), Advice-Seeking (SR = 2.83), and Advice-Offering (SR = 2.08).

Sentiment also differed across periods. Pre-COVID entries had 44.6% positive, 43.9% negative, and 11.5% neutral sentiment, with a median score of 0.00 [−0.66, 0.58]. Post-COVID sentiment saw an increase in positive sentiment, with entries being 52.4% positive, 36.9% negative, and 10.7% neutral, with a median score of 0.17 [−0.46 to 0.72].

A Mann-Whitney U test confirmed a statistically significant increase in sentiment scores post-COVID ($W = 384615$, $P = .01$), suggesting a shift toward a more positive tone in HPV vaccine discussions since the onset of the pandemic.

When sentiment changes were assessed within individual themes, no category-specific comparisons reached statistical significance (all $P > .05$). However, some themes demonstrated notable shifts. For example, vaccine hesitancy moved from a mean of −0.27 to +0.09 ($P = .09$), and social or cultural influences shifted from −0.17 to +0.14 ($P = .08$). While not statistically significant, these trends may indicate evolving tone and engagement, possibly reflecting more constructive or reflective dialogue in the post-COVID landscape.

Linear Regression of Themes Predicting Sentiment Scores

The linear regression model examining theme as a predictor of sentiment score was not statistically significant ($F(9, 2791) = 0.98$, $P = .45$), indicating that thematic category did not significantly predict

sentiment score ($R^2 = 0.003$) when controlling for other themes using "Factual Claims" as the reference category. No individual theme coefficient was statistically significant.

Median Response Time

To assess user interaction and determine whether Reddit serves as a reliable platform for receiving timely responses to vaccine-related inquiries, we analyzed the median time from post to first comment. The overall median response time was 136.1 minutes, suggesting that users typically received replies within a few hours. However, engagement varied by subreddit. Among the most active communities, r/HPV (850 unique posts) had a median response time of 148.2 minutes, r/PreCervicalCancer (110 posts) had a response time of 215.6 minutes, and r/VACCINES (102 posts) showed a faster response time of 109.8 minutes. Additional subreddits included r/Vaccine (29 posts) with a median time of 310.1 minutes, r/CervicalCancer (18 posts) at 94.1 minutes, r/DebateVaccines (17 posts) at 115.7 minutes, r/vaxxhappened (13 posts) at 417.9 minutes, and r/AskDocs (10 posts) at 77.6 minutes. Subreddits with fewer than 10 unique posts are summarized in Table 1 and were excluded from direct comparisons due to small sample size. This response time analysis, which included all collected entries (both intentional and incidental) for posts initiating threads, suggests Reddit is an interactive platform where users can expect responses to posts within a relatively short timeframe.

Discussion

This study utilized manual thematic coding and sentiment analysis on 2,801 Reddit entries about the HPV vaccine to map dominant discussion topics, sentiment trends, and shifts before versus after the COVID-19 pandemic.

Factual claims (25.1 %), personal experiences (20.4 %), and advice-offering (16.6 %) were the most frequent themes (Table 3), underscoring Reddit's dual role as both an information repository and a peer-support forum for health-related decision-making. With an overall median post-to-first-comment time of just over 2 hours and subreddit communities with member counts in the tens or hundreds of thousands, users can quickly ask questions and receive feedback. This responsiveness is particularly relevant in the post-COVID era, during which patient portal messages increased by 157%, potentially causing delays in direct communication with healthcare providers.²² The marked rise in HPV vaccine discussions after 2020 (Figure 2) may therefore reflect a growing demand for timely, community-driven health information amid pandemic-related uncertainty.

Post-pandemic discussions shifted away from systemic concerns (e.g., access barriers, social/cultural influences) toward more personal and advice-oriented posts. This evolution, confirmed by a significant χ^2 test of theme distributions, may reflect users' growing comfort sharing lived experiences and peer recommendations in a more supportive online environment.

We observed a significant post-COVID uptick in positive sentiment (Mann-Whitney U, $P=.01$), which may reflect increased comfort with discussing vaccines online or greater reliance on peer-shared health information during the pandemic.

While some studies report a decline in vaccine confidence post-COVID-19,²³ a recent literature review found increased uptake of routine adult vaccinations, such as influenza and pneumococcal vaccines, alongside decreased childhood vaccination rates.²⁴ Additionally, a cross-sectional survey showed that most participants had no change in opinion regarding the HPV vaccine following the pandemic.²⁵ These contrasting findings underscore the complex and evolving nature of public

attitudes toward vaccination and suggest that confidence may vary by population, vaccine type, and perceived risk.

The shift toward more experience-driven content and increasingly positive sentiment may point to Reddit's growing role as an informal support network and a disseminator of health information. Users appear to use the platform not just to access information, but also to validate their concerns, share lived experiences, and support others navigating vaccine decisions. This trend is further reflected in the growing volume of HPV vaccine-related posts over time. Poisson regression modeling showed a 50% annual increase in post volume from 2016 to 2025, with a sharper rise post-COVID (60% per year) compared to pre-COVID (13% per year), all statistically significant. Public health professionals might consider monitoring Reddit and similar platforms to identify emerging concerns, counter misinformation, and tailor outreach strategies to align with community-driven dialogue.

However, the unmoderated nature of many social media discussions allows misinformation to spread rapidly. For instance, a prior 2021 study identified misinformation in 25.6% of HPV vaccine-related Reddit posts.²¹ These falsehoods can complicate public health messaging and create confusion among users. More studies similar to the work of Du et al. should explore subreddit-specific misinformation trends to better understand the nature of these inaccuracies and to support the development of targeted interventions.²¹ For instance, Reddit users often frame their concerns in personal or culturally specific terms. Public health campaigns may benefit from tailored messaging that acknowledges these concerns and shares relatable narratives or culturally competent information. Interventions might focus especially on younger adults, who make up a large proportion of Reddit users and who may still be within vaccine-eligible age groups.

Another notable finding was the shift in sentiment within the vaccine hesitancy and social/cultural influence themes. Although not statistically significant, the directional trend may still hold public health relevance. Both categories shifted from negative sentiment pre-COVID to positive sentiment post-COVID, resulting in an overall neutral average. Although individual posts were not examined in detail, this shift may suggest a transition from overt skepticism to more reflective or constructive dialogue. For example, users may have become more likely to express uncertainty rather than opposition, share stories about overcoming hesitancy, or describe how they navigated familial or cultural pressures. These evolving dynamics within Reddit HPV discussions suggest a possible shift toward more open and solution-oriented conversations. While this trend warrants further investigation across other platforms, it represents a meaningful opportunity for targeted public health messaging.

Interestingly, the proportion of positive sentiment (51.7%) approximates recent U.S. HPV vaccination rates, with 66% of adolescents completing the full vaccine series by age 17.⁶ While demographic data were not collected in this study, prior research shows that Reddit users are predominantly under 35, with one study citing an average user age of 25.7, and most reside in English-speaking countries.^{21,26} This demographic overlap strengthens the case for using Reddit as a window into the attitudes of vaccine-eligible populations and underscores its value in identifying motivators or barriers to vaccine uptake.

Our findings expand on prior research examining HPV vaccine discourse online. Lama et al. analyzed Reddit posts from 2007–2015 and identified dominant topics such as political debate, sexual behavior, and vaccination schedules using automated topic modeling.¹⁹ While their study provided early insights into Reddit-based vaccine discussions, it did not assess sentiment or examine

temporal shifts in discourse. In addition, Du et al. focused specifically on identifying misinformation related to the HPV vaccine on Reddit, offering insight into the prevalence of inaccurate claims within public discourse, but without thematic categorization or sentiment analysis.²¹

More recently, Meci et al. examined sentiment in Reddit posts but limited their analysis to only three broad themes—factual, personal, and conspiracy—and excluded user comments.²⁰ Their study focused on discourse surrounding FDA guidance encouraging HPV vaccination for head and neck cancer prevention. In contrast, the present study analyzed both posts and comments, employed manual coding across 10 distinct themes for greater granularity, and specifically investigated temporal shifts in theme distribution and sentiment surrounding the COVID-19 pandemic. This multifaceted approach offers a more nuanced and contemporary understanding of online vaccine discourse.

Overall, our study complements prior literature by offering a more recent and detailed analysis of HPV vaccine discourse on Reddit, incorporating ten distinct themes to better capture the nuances of public discussion. Additionally, we analyzed sentiment and uniquely examined how sentiment and thematic focus shifted over time relative to the COVID-19 pandemic. Application of manual labeling enhanced contextual accuracy, especially for entries that referenced the vaccine tangentially but were not primarily vaccine-focused (e.g., wart management or cancer treatment). While this process is more time-intensive than automated methods, it improves data quality, especially when entries are long, nuanced, or touch on multiple overlapping ideas. Additionally, it eliminates bot-generated content, enhancing the reliability of our findings.

Limitations

This study is not without inherent limitations. First, the data originated from user-generated content on Reddit, introducing potential selection bias as Reddit users are not representative of the general population in terms of demographics and viewpoints. Reporting bias may also exist in how users frame their experiences or opinions. While prior research suggests that Reddit users are predominantly under the age of 35 and primarily located in English-speaking countries, the platform's anonymity makes it difficult to draw definitive conclusions about user demographics.^{21,26} Additionally, although our analysis focused on English-language content, many posts referenced non-U.S. healthcare systems, potentially limiting generalizability to U.S. populations. However, we did not exclude international posts, as the goal was to understand broader public discourse, particularly on platforms not constrained by geographic borders.

Furthermore, potential data limitations include the archiving of older posts by some subreddits and constraints of the Reddit API, possibly contributing to lower retrieved entry counts in the initial years of the study period (2016–2019). Assigning each entry to a single dominant theme, while necessary for categorization, may oversimplify posts or comments containing multiple thematic elements. Although the use of two independent coders resolving discrepancies by consensus aimed to maximize consistency and reliability, some nuance may inevitably be lost in single-theme assignment.

Lastly, the reliance on an automated tool (VADER) for sentiment analysis presents limitations. While efficient for large datasets, automated scoring may fail to capture sarcasm, context-dependent nuances, or complex emotional expressions accurately.

Future Directions

Future research could employ more advanced NLP (natural language processing) techniques (e.g., topic modeling within themes, stance detection) to further dissect nuanced discussion threads. This could enable a more fine-grained identification of specific concerns, misinformation patterns, or information gaps within each theme. In turn, such insights can better inform public health campaigns and digital outreach efforts aimed at improving vaccine uptake, trust, and health literacy in online spaces.

Additionally, future studies could explore whether awareness of HPV-related oropharyngeal cancers, particularly in men, is increasing over time through social media discourse. Prior survey research has documented low public awareness of the link between HPV and oropharyngeal cancer in men.²⁷ Because HPV is often associated with cervical cancer, many men may not perceive the vaccine as relevant to them, contributing to lower uptake and greater vaccine hesitancy in this group. A broader assessment of these knowledge gaps across social platforms could support the development of targeted, gender-specific educational campaigns. Tailored messaging that addresses misconceptions and highlights the relevance of HPV vaccination for men may be critical to improving coverage and informed decision-making in vaccine-hesitant populations.

Conclusion

This study provides a contemporary analysis of HPV vaccine discourse and sentiment on Reddit, highlighting significant shifts in both thematic focus and sentiment before and after the COVID-19 pandemic. Findings suggest that conversations have become increasingly personal and advice-driven over time, with a notable rise in positive sentiment and public engagement following 2020. These trends reflect a growing reliance on online communities for health-related dialogue and peer support, particularly within a digital landscape transformed by the pandemic.

Reddit appears to serve not only as a platform for information exchange, but also as a space for collective decision-making, emotional validation, and rapid feedback—especially in the context of healthcare system delays and uncertainty. These insights underscore the importance of public health professionals monitoring and potentially engaging with evolving online conversations to promote accurate vaccine information and respond to emerging concerns in real time.

This work builds upon prior research while offering new perspectives into the dynamic, community-driven nature of digital health communication in the social media era.

Acknowledgements

The authors acknowledge the use of artificial intelligence (AI) tools in the development of the Python script used for data extraction and the R code used for statistical analysis. AI assistance was limited to code generation and debugging, with all outputs reviewed, tested, and validated by the authors to ensure accuracy and appropriateness for the study.

Funding

No external funding was received for the conduct of this study. The authors acknowledge that the applicable article processing fee will be covered by the Department of Otolaryngology–Head and Neck Surgery at Rush University Medical Center, should the manuscript be accepted for publication.

Conflicts of Interest

None declared.

Data Availability

The dataset generated and analyzed during this study will be openly available in *JMIR Data* at the time of publication.

Authors' Contributions

Conceptualization: MM, SAS, MKB

Data curation: MM, SAS

Formal analysis: MM, SAS

Investigation: MM, SAS, DB

Methodology: MM, SAS

Project administration: MM, SAS, MKB

Software: SAS

Supervision: MKB

Validation: MM, SAS, MKB

Visualization: MM

Writing – original draft: MM, SAS, DB, JM, RP

Writing – review & editing: EMS, VCS, MKB, JM, RP

Abbreviations

API - Application Programming Interface

HPV – Human Papillomavirus

SR – Standardized Residual

U.S. - United States

VADER - Valence Aware Dictionary and Sentiment Reasoner

References

1. Szymonowicz KA, Chen J. Biological and clinical aspects of HPV-related cancers. *Cancer Biol Med.* 2020;17(4):864-878. doi:10.20892/j.issn.2095-3941.2020.0370
2. Chabeda A, Yanez RJR, Lamprecht R, Meyers AE, Rybicki EP, Hitzeroth II. Therapeutic vaccines for high-risk HPV-associated diseases. *Papillomavirus Res.* 2018;5:46-58. doi:10.1016/j.pvr.2017.12.006
3. Egawa N, Doorbar J. The low-risk papillomaviruses. *Virus Res.* 2017;231:119-127. doi:10.1016/j.virusres.2016.12.017
4. Ljubojević S. The human papillomavirus vaccines. *Acta Dermatovenerol Croat ADC.* 2006;14(3):208.
5. Harper DM, DeMars LR. HPV vaccines – A review of the first decade. *Gynecol Oncol.* 2017;146(1):196-204. doi:10.1016/j.ygyno.2017.04.004
6. Pingali C. National Vaccination Coverage Among Adolescents Aged 13–17 Years — National

- Immunization Survey-Teen, United States, 2023. *MMWR Morb Mortal Wkly Rep.* 2024;73. doi:10.15585/mmwr.mm7333a1
7. Cervical Cancer Elimination Initiative. Accessed January 29, 2025. <https://www.who.int/initiatives/cervical-cancer-elimination-initiative>
 8. Saulsberry L, Fowler EF, Nagler RH, Gollust SE. Perceptions of politicization and HPV vaccine policy support. *Vaccine.* 2019;37(35):5121-5128. doi:10.1016/j.vaccine.2019.05.062
 9. Kornides ML, Badlis S, Head KJ, Putt M, Cappella J, Gonzalez-Hernandez G. Exploring content of misinformation about HPV vaccine on twitter. *J Behav Med.* 2023;46(1):239-252. doi:10.1007/s10865-022-00342-1
 10. Fowler EF, Nagler RH, Banka D, Gollust SE. Chapter Six - Effects of politicized media coverage: Experimental evidence from the HPV vaccine and COVID-19. In: Bolsen T, Palm R, eds. *Progress in Molecular Biology and Translational Science.* Vol 188. Molecular Biology and Clinical Medicine in the Age of Politicization. Academic Press; 2022:101-134. doi:10.1016/bs.pmbts.2021.11.008
 11. Du J, Luo C, Shegog R, et al. Use of Deep Learning to Analyze Social Media Discussions About the Human Papillomavirus Vaccine. *JAMA Netw Open.* 2020;3(11):e2022025. doi:10.1001/jamanetworkopen.2020.22025
 12. Shapiro GK, Surian D, Dunn AG, Perry R, Kelaher M. Comparing human papillomavirus vaccine concerns on Twitter: a cross-sectional study of users in Australia, Canada and the UK. *BMJ Open.* 2017;7(10):e016869. doi:10.1136/bmjopen-2017-016869
 13. Du J, Xu J, Song HY, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. *BMC Med Inform Decis Mak.* 2017;17(Suppl 2):69. doi:10.1186/s12911-017-0469-6
 14. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. *J Med Internet Res.* 2015;17(6):e144. doi:10.2196/jmir.4343
 15. Surian D, Nguyen DQ, Kennedy G, Johnson M, Coiera E, Dunn AG. Characterizing Twitter Discussions About HPV Vaccines Using Topic Modeling and Community Detection. *J Med Internet Res.* 2016;18(8):e232. doi:10.2196/jmir.6045
 16. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, 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. doi:10.1016/j.vaccine.2017.04.060
 17. Zhou X, Coiera E, Tsafnat G, Arachi D, Ong MS, Dunn AG. Using social connection information to improve opinion mining: Identifying negative sentiment about HPV vaccines on Twitter. *Stud Health Technol Inform.* 2015;216:761-765.
 18. Luo X, Zimet G, Shah S. A natural language processing framework to analyse the opinions on HPV vaccination reflected in twitter over 10 years (2008 - 2017). *Hum Vaccines Immunother.* 2019;15(7-8):1496-1504. doi:10.1080/21645515.2019.1627821

19. Lama Y, Hu D, Jamison A, Quinn SC, Broniatowski DA. Characterizing Trends in Human Papillomavirus Vaccine Discourse on Reddit (2007-2015): An Observational Study. *JMIR Public Health Surveill.* 2019;5(1):e12480. doi:10.2196/12480
20. Meci A, Tseng CC, Jensen Z, Goyal N. Online Discourse and Trends Surrounding HPV Vaccination for Head and Neck Cancer Prevention. *Otolaryngol Neck Surg.* 2025;n/a(n/a). doi:10.1002/ohn.1222
21. Du J, Preston S, Sun H, et al. Using Machine Learning–Based Approaches for the Detection and Classification of Human Papillomavirus Vaccine Misinformation: Infodemiology Study of Reddit Discussions. *J Med Internet Res.* 2021;23(8):e26478. doi:10.2196/26478
22. Holmgren AJ, Downing NL, Tang M, Sharp C, Longhurst C, Huckman RS. Assessing the impact of the COVID-19 pandemic on clinician ambulatory electronic health record use. *J Am Med Inform Assoc.* 2022;29(3):453-460. doi:10.1093/jamia/ocab268
23. Siani A, Tranter A. Is vaccine confidence an unexpected victim of the COVID-19 pandemic? *Vaccine.* 2022;40(50):7262-7269. doi:10.1016/j.vaccine.2022.10.061
24. Wojtek I, Wouw M van, Thomson A. Impact of COVID-19 on vaccine confidence and uptake: A systematic literature review. *Hum Vaccines Immunother.* Published online December 31, 2024. Accessed April 24, 2025. <https://www.tandfonline.com/doi/abs/10.1080/21645515.2024.2384180>
25. Schelbar N, Ward CN, Phillips E, et al. Impact of COVID-19 pandemic and vaccine perceptions on HPV vaccine hesitancy. *Am J Otolaryngol.* 2024;45(2):104172. doi:10.1016/j.amjoto.2023.104172
26. Gjurković M, Karan M, Vukojević I, Bošnjak M, Šnajder J. PANDORA Talks: Personality and Demographics on Reddit. Published online June 8, 2021. doi:10.48550/arXiv.2004.04460
27. Davis EN, Doyle PC. An Assessment of Young Adults' Awareness and Knowledge Related to the Human Papillomavirus (HPV), Oropharyngeal Cancer, and the HPV Vaccine. *Cancers.* 2025;17(3):344. doi:10.3390/cancers17030344

Tables

Table 1. Data Sources: Subreddit Communities Included and Applied Keyword Filters.

Subreddit Category	Subreddits	Keywords Applied
HPV-Focused	r/HPV, r/PreCervicalCancer, r/Warts, r/CervicalCancer	Vaccine-Related Keywords: - vaccine, vaccines, vaccinated, vaccination, vaccinations, vax, vaxxed, unvaxxed - Gardasil, cervarix - anti-vax, antivax, antivaccine, anti-vaccine, pro-vax - jab, shot, immunization
Vaccine-Focused	r/VACCINES, r/vaxxhappened, r/Vaccine, r/antivax, r/DebateVaccines, r/unvaccinated, r/skeptic, r/StopMandatoryVaccine	HPV-Related Keywords: - HPV, HPV vaccine, HPV shot - gardasil, cervarix - human papillomavirus

Medical & Science	r/AskDocs, r/medicine, r/PeterAttia, r/HeadandNeckCancer, r/Health, r/ScienceBasedParenting, r/HNSCC	HPV + Vaccine-Related Keywords: - HPV, HPV vaccine, HPV shot, gardasil, cervarix, human papillomavirus - vaccine, vaccines, vaccinated, vaccination, jab, shot, immunization - anti-vax, antivax, pro-vax
Women's Health	r/TwoXChromosomes, r/TheGirlSurvivalGuide, r/WomenOver40, r/AskWomenOver30	
General & Social	r/MensRights, r/AskGaybrosOver30, r/AskConservatives, r/askgaybros	

Table 2. Thematic Codebook: Labels, Definitions, and Sample Excerpts.

Theme	Definition	Examples
Advice/Recommendations (Offering)	Posts that explicitly offer advice or recommendations regarding HPV vaccination to others.	"You should get vaccinated", "My doctor recommended getting the vaccine"
Advice/Recommendations (Seeking)	Posts that request advice or opinions from others about receiving the HPV vaccine.	"Should I get the vaccine?", "Should I encourage my partner to get the vaccine?"
Barriers/Accessibility	Posts that describe challenges in accessing the HPV vaccine, including financial, insurance, geographic, or age-related limitations.	"I wanted the vaccine 3 months before I was 26, insurance said no, my age disqualified coverage, but I hadn't turned 26 yet", "My daughter got the shot after she moved out and the first shot was over 400 dollars. I told her about in network and the 2nd was free."
Factual Claims	Statements presented as facts, regardless of accuracy, typically lacking overt opinion or emotion. Intended to convey information or assert knowledge.	"Gardasil 9 protects against nine types of human papillomavirus (HPV): 6, 11, 16, 18, 31, 33, 45, 52, and 58", "There's zero data on actual cervical cancer rates declining due to HPV vaccines."
General Questions	Posts that inquire about specific aspects of the HPV vaccine, including dosing, efficacy, safety, or eligibility.	"Is two shots enough to protect you?", "Why the difference between US CDC 3-dose and UK NHS 2-dose recommendation for HPV vaccination for older adults up to age 45?"
Hesitancy	Posts expressing concerns, skepticism, or fear about HPV vaccine safety, efficacy, or perceived risks. May include conspiracy or mistrust themes.	"I know someone who periodically needs to have her skull drained of excess fluid because of the HPV vaccine. They almost died after taking it due to fluid build up around her brains. I stupidly took it as an adult at 36 and have had

		autoimmune issues ever since.”, “And my generation, 90% of us have HPV and don't know it and we're fine... U think doctors don't lie? Doctors aren't corrupt or capable of being corrupted? Just use common sense. Less vaccines the better IMHO!”
Personal Experience	Narratives sharing personal or family experiences with HPV vaccination, including positive, negative, or neutral outcomes.	“I got my 1st dose of Gardasil vaccine on Thursday. I have a fever of 103.3 and I feel really sick. How are you doing?”, “I was diagnosed with HPV after almost 20 years of marriage... I'm now getting vaccinated at 43, even though I had a hysterectomy because I want to make sure I'm protected from any possible further in the future.”, “I had the vaccine aged 12, and got diagnosed with HPV at my first smear, ended up having a colposcopy to remove dodgy cells. It is not a guaranteed prevention.”
Resources/Data Sharing	Posts that share articles, external links, or statistical information to support a claim, argument, or question.	“The CDC recommends that women get the HPV vaccine through age 26, if they haven't already been vaccinated. https://www.cancer.gov/types/cervical/causes-risk-prevention ”, “The risk of cancer for low risk strains is quite low. The high risk strains themselves comprise the majority of HPV-derived cancer. See [this table](https://www.cdc.gov/cancer/hpv/statistics/cases.htm) for a breakdown of risk of cancer type by HPV strain.”
Social/Cultural Influences	Posts discussing how culture, religion, family, or societal norms influence attitudes or decisions regarding HPV vaccination.	“I thought, because of what the Ministry of Health informed back then in my country, the HPV vaccine only worked in young girls.”, “I never got the vaccine when I was younger (due to being too young to make my own decision and my mom being against it). I'm obviously very upset but just happy to see this post and hopefully my HPV clears up and goes dormant once this is all over with”
Other	Posts that don't fit into other categories due to ambiguity, sarcasm, or lack of context.	“At what age did you get the vaccine?”, “I'm not her, so no - her witness testimony is evidence of her claims & she's more credible than Gardasil.”

Table 3. Theme Frequencies and Mean Sentiment Scores Before versus After World Health Organization's COVID-19 Pandemic Declaration (March 11, 2020).

Theme	Pre-COVID Count (%)	Pre-COVID Sentiment (mean ± SD)	Post-COVID Count (%)	Post-COVID Sentiment (mean ± SD)	Total Count (%)	Total Sentiment (mean ± SD)
Advice/ Recommendations (Offering)	34 (12.2%)	0.05 ± 0.69	432 (17.1%)	0.12 ± 0.63	466 (16.6%)	0.12 ± 0.64
Advice/ Recommendations (Seeking)	5 (1.8%)	-0.09 ± 0.30	148 (5.9%)	0.18 ± 0.63	153 (5.5%)	0.17 ± 0.62
Barriers/ Accessibility	52 (18.7%)	0.03 ± 0.71	168 (6.7%)	0.12 ± 0.63	220 (7.8%)	0.10 ± 0.65
Factual Claims	71 (25.5%)	0.00 ± 0.62	632 (25.0%)	0.09 ± 0.65	703 (25.1%)	0.08 ± 0.65

General Questions	16 (5.8%)	-0.01 ± 0.56	199 (7.9%)	0.08 ± 0.69	215 (7.7%)	0.07 ± 0.68
Hesitancy	9 (3.2%)	-0.27 ± 0.52	132 (5.2%)	0.09 ± 0.65	141 (5.0%)	0.07 ± 0.64
Personal Experience	37 (13.3%)	0.08 ± 0.62	536 (21.2%)	0.14 ± 0.64	573 (20.4%)	0.13 ± 0.64
Resources/Data Sharing	21 (7.6%)	0.05 ± 0.70	161 (6.4%)	0.10 ± 0.66	182 (6.5%)	0.09 ± 0.67
Social/Cultural Influences	28 (10.1%)	-0.17 ± 0.73	62 (2.5%)	0.14 ± 0.68	90 (3.2%)	0.04 ± 0.71
Other	5 (1.8%)	0.11 ± 0.67	53 (2.1%)	-0.05 ± 0.65	58 (2.1%)	-0.04 ± 0.65
Total	278	0.00 ± 0.65	2,523	0.11 ± 0.65	2,801	0.10 ± 0.65

Note. Counts represent the total number of entries per theme in each period. Mean sentiment scores (± SD) are shown alongside their respective theme frequencies.

Figures

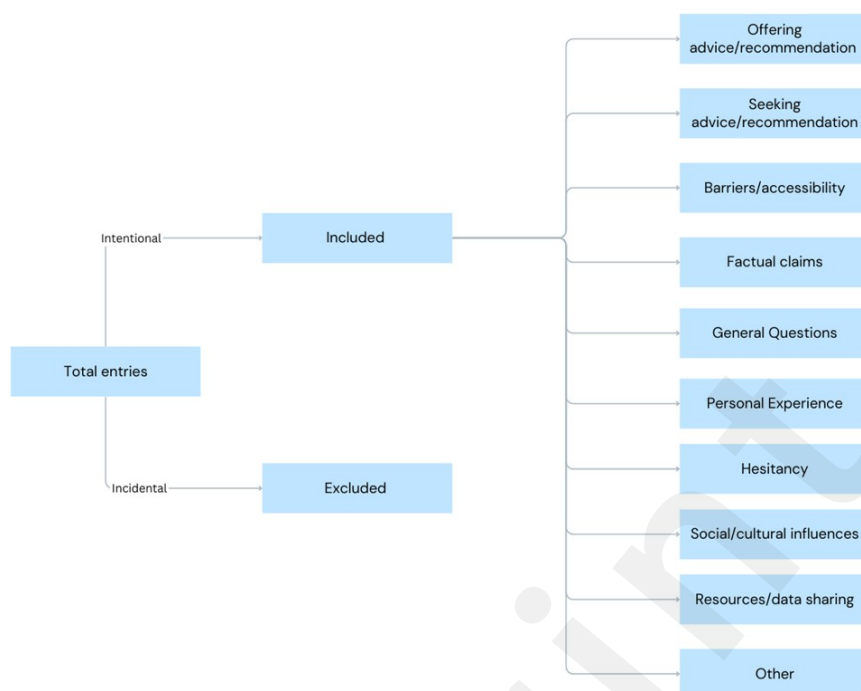


Figure 1. Flowchart of Post and Comment Screening and Thematic Coding. Reddit entries were first screened for relevance and classified as Intentional (explicitly addressing HPV vaccination) or Incidental (unrelated content, excluded). Intentional posts were then subjected to thematic coding into ten mutually exclusive categories.

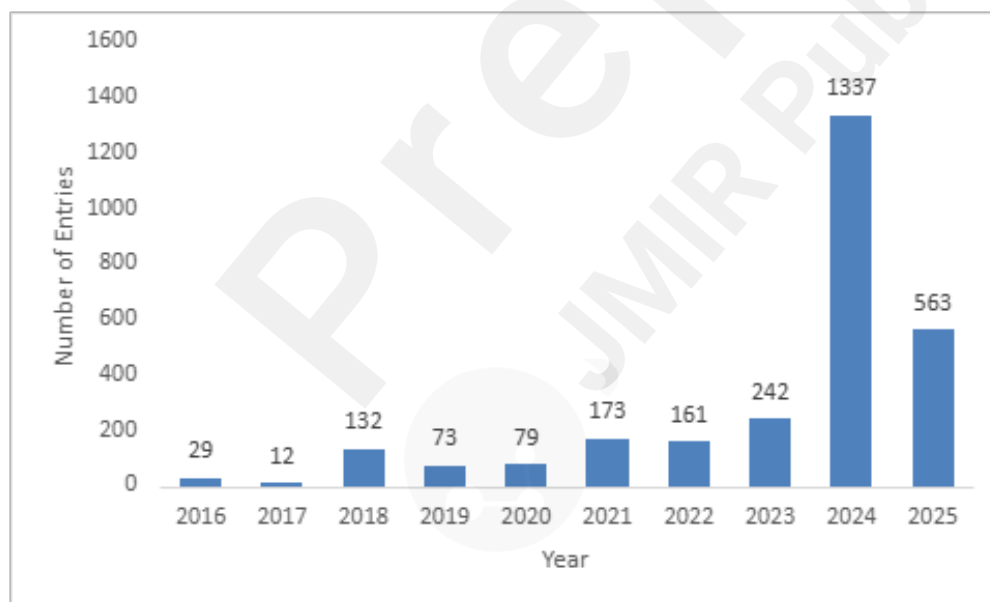
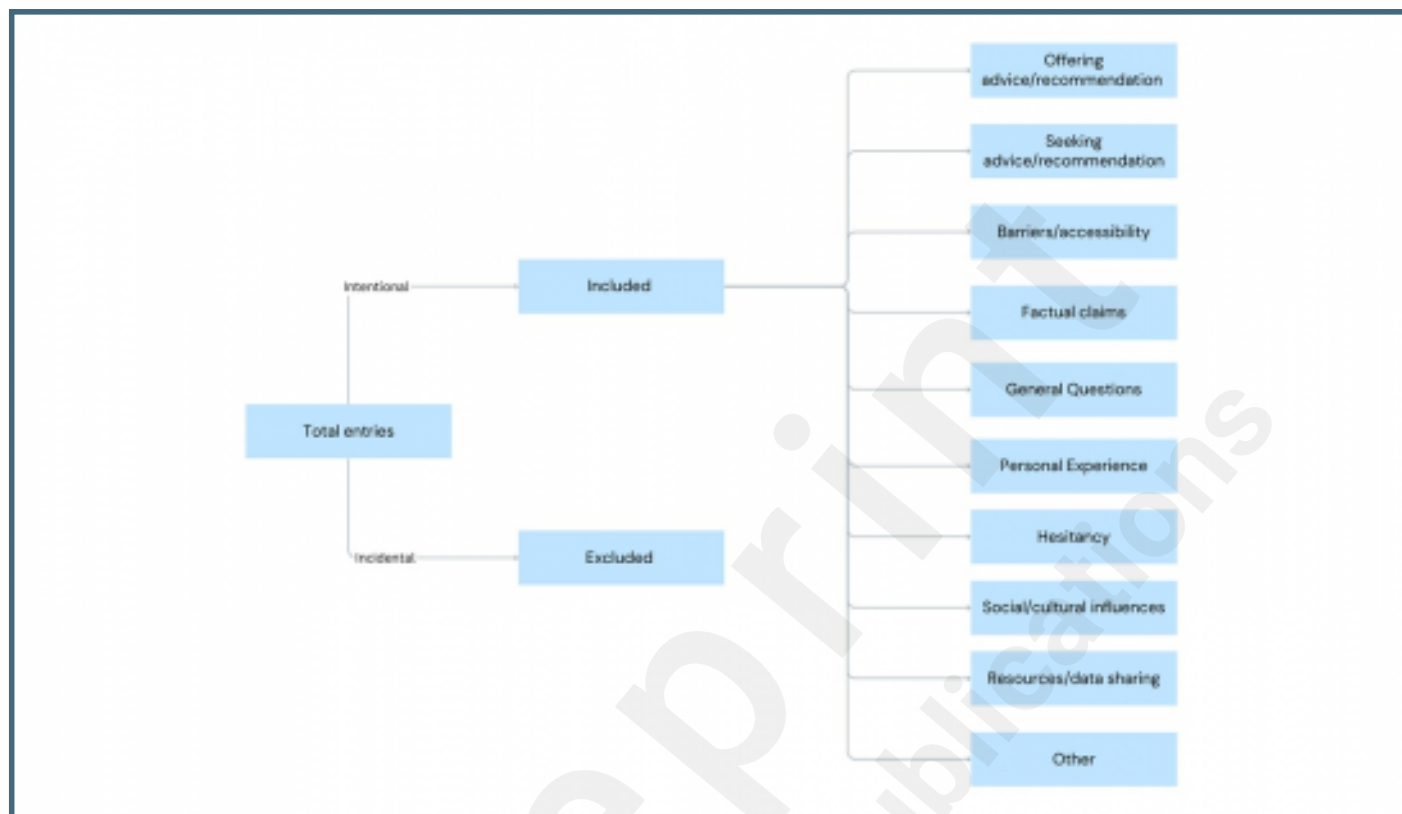


Figure 2. Annual Volume of Reddit Posts Referencing the HPV Vaccine, September 2016–February 2025. Bars indicate the total number of Reddit entries containing HPV vaccine-related keywords in each calendar year.

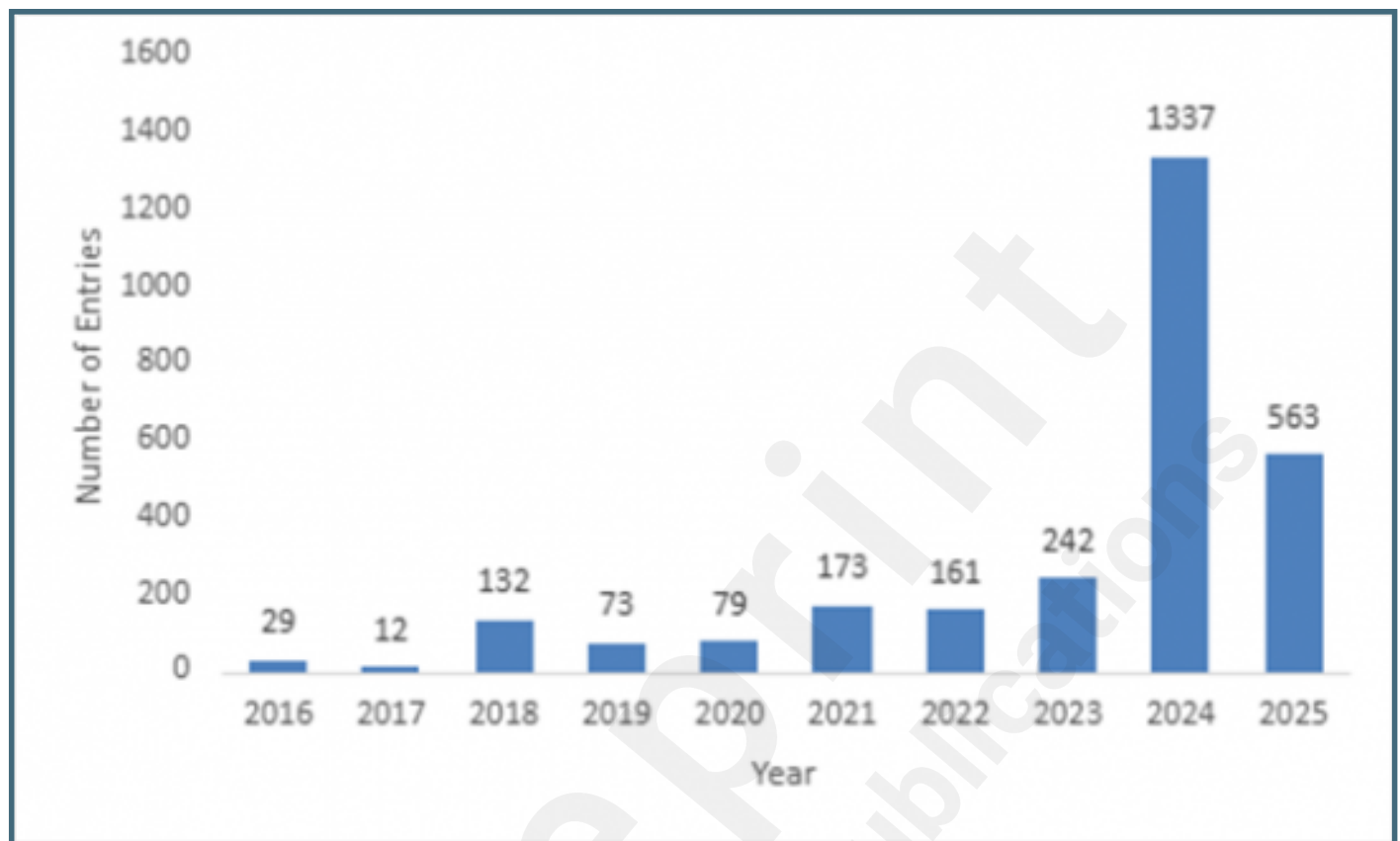
Supplementary Files

Figures

Flowchart of post and comment screening and thematic coding. Reddit entries were first screened for relevance and classified as Intentional (explicitly addressing HPV vaccination) or Incidental (unrelated content, excluded). Intentional posts were then subjected to thematic coding into ten mutually exclusive categories.



Annual volume of reddit posts referencing the HPV vaccine, September 2016– February 2025. Bars indicate the total number of Reddit entries containing HPV vaccine–related keywords in each calendar year.



Multimedia Appendixes

Microsoft Excel spreadsheet containing all posts and comments scraped, along with their labels and sentiment scores.

URL: <http://asset.jmir.pub/assets/09da4360b23b2550fd17adbb6363e81c.xlsx>

Natural language processing code used for sentiment analysis.

URL: <http://asset.jmir.pub/assets/558d234eb286d2852308d75ac2c890ab.pdf>

Python script used to scrape data from Reddit.

URL: <http://asset.jmir.pub/assets/e38058907f44f6ba167f99b7d7435bcb.pdf>



Related publication(s) - for reviewers eyes onlies

Untitled.

URL: <http://asset.jmir.pub/assets/5250514ae3f972ea306a5ffafba51f9a.pdf>

Untitled.

URL: <http://asset.jmir.pub/assets/10c7a8d1f49dcd78f301c7fdb41a7fda.pdf>