Engagement With HPV-Related Content on Instagram: A Cross-Sectional Descriptive and Comparative Analysis of Posts Retrieved Through Four Hashtags

Cem Yagmur Özdemir[1,*], Aykut Şimşek[2], Nayif Çiçekli[3] and Nagihan Özdemir[4]

[1]*Department of Gynecological Oncology, Afyonkarahisar State Hospital, Afyonkarahisar, Turkey*

[2]*Meta Platforms, Software Engineering Manager, London, England*

[3]*Department of Gynecological Oncology, Erzurum City Hospital, Erzurum, Turkey*

[4]*Department of Dermatology, Faculty of Medicine, Afyonkarahisar Health Sciences University, Afyonkarahisar, Turkey*

Abstract: This cross-sectional descriptive and comparative study described HPV-related Instagram posts retrieved through four predefined hashtags and examined associations of current follower count, account type, caption language, and post age with like and comment counts. Publicly available posts published between March 5, 2014, and September 15, 2025, were retrieved on October 15, 2025, using Instaloader. After non-post records and duplicates were removed, 1,281 unique posts from 685 accounts were included. Negative binomial generalized estimating equation models accounted for clustering within accounts and adjusted for current follower count, post age, account type, and caption language. Account-clustered negative binomial regression was used as a sensitivity analysis. Median engagement was 3 likes (IQR, 0 to 16) and 0 comments (IQR, 0 to 1). Each doubling of current follower count was associated with higher like counts (adjusted incidence rate ratio [aIRR], 1.43; 95% CI, 1.37 to 1.50) and comment counts (aIRR, 1.43; 95% CI, 1.32 to 1.54). Each doubling of post age was also associated with higher like counts (aIRR, 1.33; 95% CI, 1.14 to 1.56) and comment counts (aIRR, 1.39; 95% CI, 1.18 to 1.64). Influencer or advocacy accounts did not differ significantly from healthcare professional accounts after adjustment, and caption language was not associated with either outcome. Sensitivity analyses were consistent. Current follower count and post age were associated with accumulated engagement, but these findings should not be interpreted as evidence of content quality, professional authority, historical visibility, or direct public health impact.

Keywords: Papillomavirus Infections, Social Media, Health Communication, Online Social Networking, Information Dissemination, Uterine Cervical Neoplasms.

INTRODUCTION

Human papillomavirus (HPV) is one of the most common sexually transmitted infections worldwide and is causally associated with cervical cancer and several other anogenital malignancies [1,2]. Although effective vaccination and screening strategies are available, HPV related diseases continue to impose a substantial global health burden, particularly in settings with suboptimal vaccination coverage and limited access to organized screening programs [3,4]. Public awareness and clear, evidence based health communication are therefore important for supporting preventive behaviors, including HPV vaccination and participation in cervical cancer screening. Social media platforms are increasingly used to access, exchange, and disseminate health information, particularly among younger populations. Instagram is relevant to health communication because its visual format, hashtag based content discovery, and publicly visible interaction metrics facilitate the rapid circulation of health related messages [5]. These features may support awareness and peer communication, but they may also contribute to the dissemination of incomplete, non evidence based, or misleading claims [6,7]. In the context of HPV, social media content may shape perceptions of infection, vaccination, stigma, and prevention related behaviors [8,9]. However, platform engagement indicators such as likes and comments reflect interaction with content and should not be interpreted as direct evidence of improved knowledge, vaccine uptake, screening participation, or other public health outcomes.

Previous research on HPV-related communication through social media has primarily examined content characteristics, misinformation, vaccination-related messaging, and potential influences on awareness or vaccine attitudes, with much of the available evidence derived from platforms other than Instagram [10-13]. Evidence specifically examining engagement with HPV-related Instagram posts remains limited. In particular, few studies have evaluated like and comment counts in relation to account-level characteristics while accounting for repeated posts

*Address correspondence to this author at the Department of Gynecological Oncology, Afyonkarahisar State Hospital, Orhangazi, Afyonkarahisar State Hospital, 03030 Center, Afyonkarahisar, Turkey;
E-mail: cyozdemir@hotmail.com

ISSN: 1929-2260 / E-ISSN: 1929-2279/26

from the same account and the unequal periods during which posts accumulate engagement. Furthermore, current follower count should be regarded as a contemporaneously measured account characteristic rather than a measure of a post's historical visibility, account type should be treated as a descriptive classification rather than a direct measure of professional authority, and caption language should not be interpreted as a proxy for geographic location.

Accordingly, the present study aimed to describe HPV-related Instagram posts retrieved through four predefined hashtags and to examine the associations of current follower count, account type, caption language, and post age with like and comment counts. The analyses accounted for repeated posts originating from the same account and for differences in the time available for posts to accumulate engagement. The study was intended to characterize engagement patterns within the retrieved sample rather than to assess content quality, professional authority, geographic distribution, or direct public health impact.

METHODS

Study Design and Data Source

This cross-sectional descriptive and comparative study analyzed publicly available Instagram posts related to human papillomavirus. Instagram was selected because its hashtag-based content organization, visual communication format, and publicly visible post-level interaction metrics allowed standardized retrieval and comparison of HPV-related posts. The study was designed specifically to characterize engagement patterns on Instagram and was not intended as a comparison across social media platforms. Data were retrieved on October 15, 2025, using Instaloader (https://instaloader.github.io), an open-source Python-based software package developed by Alexander Graf and André Koch-Kramer in Germany and first released in 2016. Instaloader enables the retrieval of publicly accessible Instagram posts and associated metadata, including publication date, caption, account information, follower count, likes, and comments. Only publicly accessible posts and metadata were collected. No attempt was made to access private, restricted, or otherwise inaccessible content.

Post Selection and Sampling Strategy

Posts published between March 5, 2014, and September 15, 2025, were eligible for inclusion. Four hashtags were predefined before data retrieval: #hpv, #hpvawareness, #hpvsupport, and #cervicalcancer. These hashtags were selected to capture general HPV content, awareness-oriented communication, peer support content, and cervical cancer-related prevention messaging. To maintain feasibility and prevent the broadest hashtag from dominating the sample, retrieval was limited to approximately 400 consecutively available posts per hashtag, depending on platform availability at the time of extraction. Posts were not manually ranked or selected according to engagement level.

Eligible records were publicly accessible Instagram posts within the predefined publication period for which the post identifier, publication date, account information, current follower count, like count, and comment count were retrievable. Private, deleted, restricted, or otherwise inaccessible posts could not be retrieved and were therefore not included. Non-post metadata records, blank records, and duplicate posts appearing under more than one hashtag were excluded. Duplicate records were identified using the unique Instagram post identifier. The initial export contained 1,340 records. After exclusion of four hashtag-level metadata records, one blank record, and 54 redundant duplicate records, the final analytic dataset comprised 1,281 unique posts from 685 distinct accounts.

The resulting sample represents a non-probability, platform-dependent snapshot of posts retrievable through the four selected hashtags on the extraction date. It was not intended to represent all HPV-related content on Instagram or content related to every HPV-associated malignancy. In particular, site-specific hashtags for oral, oropharyngeal, anal, penile, vulvar, vaginal, and other HPV-associated diseases were not searched separately.

Data Retrieval and Investigator Roles

Data retrieval was performed by two investigators. Data cleaning, duplicate identification, and preparation of the final analytic dataset were conducted by one investigator. Account type classification and caption language classification were each performed by two investigators. Records considered uncertain during the classification process were jointly reviewed by two investigators, and final decisions were reached by consensus.

Account Type Classification

Each account was classified once, and the assigned category was applied to all posts originating from that

account. Classification was based on manual review of the publicly available username, account name, profile description, and posted content. Five mutually exclusive categories were used: healthcare professional, health organization, commercial account, influencer or advocacy account, and individual user. Healthcare professional accounts explicitly identified the account holder as a physician, nurse, midwife, psychologist, or another healthcare professional. Health organizations included hospitals, clinics, professional associations, nongovernmental organizations, public health campaigns, and other formal health related institutions. Commercial accounts primarily promoted products or paid services. Influencer or advocacy accounts primarily focused on public awareness, peer support, personal advocacy, or lifestyle communication without meeting the criteria for a healthcare professional or formal health organization. Individual users were personal accounts that did not meet the criteria for the other categories.

Account type classification was performed by two investigators. Records that remained uncertain were reviewed by two investigators and resolved by consensus. Because classification was not conducted as blinded independent duplicate coding, a formal interrater reliability statistic was not calculated. Account type was treated as a descriptive account characteristic and not as a direct measure of professional authority, credibility, content accuracy, or informational quality.

Caption Language Classification

The primary language of each post caption was classified manually as English, Spanish or Portuguese, Turkish, or other or uncertain. Spanish and Portuguese were combined because of their linguistic similarity and the relatively limited number of posts in each language group. Posts containing more than one language were classified according to the predominant language of the caption. Posts without sufficient caption text, posts consisting mainly of emojis or hashtags, and posts for which the predominant language could not be determined were categorized as other or uncertain.

Caption language classification was performed by two investigators. Uncertain records were reviewed jointly, and final classifications were reached by consensus. Caption language was analyzed only as a descriptive characteristic of the retrieved posts. It was not used as a proxy for the geographic location, nationality, cultural background, or audience region of the account.

Engagement Outcomes and Exposure Variables

The primary outcomes were the number of likes and the number of comments recorded for each post on October 15, 2025. These variables were analyzed separately as count outcomes. Engagement rate was not used because follower count would have appeared both in the denominator of the outcome and as an explanatory variable, creating a risk of mathematical coupling and potentially producing an artificial inverse association.

Current follower count was recorded for each account on the data extraction date and was treated as a contemporaneous account characteristic. It was not interpreted as the number of followers present when an older post was originally published or as a direct measure of the historical visibility or reach of that post.

Post age was calculated as the number of days between the publication date of each post and October 15, 2025. This variable was included to account for differences in the time available for posts to accumulate likes and comments. Because current follower count and post age showed markedly right skewed distributions, both variables were transformed using the base 2 logarithm before multivariable analysis. Accordingly, their incidence rate ratios represent the expected change in like or comment counts associated with a doubling of the corresponding variable.

The remaining explanatory variables were account type and caption language. Account type was entered as a five category variable, with healthcare professional accounts used as the reference group. Caption language was entered as a four category variable, with English language posts used as the reference group.

Statistical Analysis

Continuous variables were summarized using the mean and standard deviation and the median and interquartile range, whereas categorical variables were presented as frequencies and percentages. Because like and comment counts were nonnegative count variables with marked right skew, excess zero values, and variance substantially exceeding the mean, Gaussian linear regression was not considered appropriate.

Associations with like and comment counts were evaluated separately using multivariable negative binomial generalized estimating equation models.

Accounts were specified as the clustering variable to account for the nonindependence of multiple posts originating from the same account, and an exchangeable working correlation structure was used. Robust sandwich standard errors were calculated. The models included current follower count, post age, account type, and caption language. Current follower count and post age were transformed using the base 2 logarithm because of their markedly right skewed distributions. Adjusted associations were reported as incidence rate ratios with 95% confidence intervals. For these continuous variables, the incidence rate ratio represents the expected multiplicative change in the outcome associated with a doubling of current follower count or post age.

Healthcare professional accounts were used as the reference category for account type, and English was used as the reference category for caption language. Variance inflation factors were examined to assess multicollinearity. No missing values were present in the variables included in the multivariable models. The engagement rate model used in the initial analysis was removed because current follower count was included both in the denominator of that outcome and as an explanatory variable.

As a sensitivity analysis, conventional negative binomial regression models were fitted using account-clustered robust standard errors. Consistency between the primary and sensitivity analyses was evaluated according to the direction, magnitude, confidence intervals, and statistical significance of the associations. All tests were two sided, and p values below 0.05 were considered statistically significant. Analyses were performed using Python version 3.13, IBM SPSS Statistics version 22 and the statsmodels package version 0.14.6.

RESULTS

A total of 1,340 records were retrieved from the four predefined hashtags. After exclusion of four hashtag level metadata records, one blank record, and 54 redundant records representing duplicate posts retrieved through more than one hashtag, 1,281 unique Instagram posts from 685 distinct accounts were included in the final analysis (Figure **1**). No missing values were present in the variables included in the multivariable models.

Accounts contributed a median of 1 post (IQR, 1 to 1), with a range of 1 to 141 posts per account. The median current follower count was 1,098 (IQR, 153 to 4,594). The median post age at data extraction was 37 days (IQR, 34 to 48), with a range of 33 to 4,242 days. Engagement measures were markedly right skewed. The median number of likes was 3 per post (IQR, 0 to 16), and the median number of comments was 0 (IQR, 0 to 1). Overall, 404 posts (31.5%) received no likes, and 893 posts (69.7%) received no comments (Table **1**).

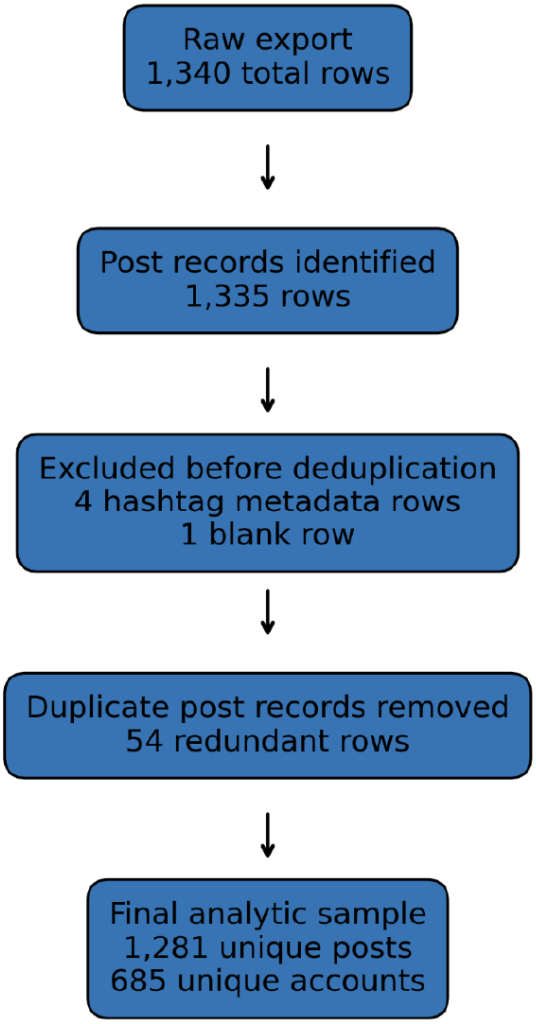

Figure 1: Flow diagram of data retrieval, cleaning, and post selection.

A total of 1,340 records were retrieved through four predefined hashtags. Four hashtag level metadata records, one blank record, and 54 redundant duplicate records were excluded, resulting in a final analytic sample of 1,281 unique Instagram posts from 685 distinct accounts.

Individual users constituted the largest account category, comprising 395 of 685 accounts (57.7%), followed by healthcare professionals (93, 13.6%), commercial accounts (74, 10.8%), influencer or advocacy accounts (62, 9.1%), and health organizations (61, 8.9%). At the post level, individual accounts contributed 482 posts (37.6%), health

Table 1: Summary of Account and Post Level Characteristics

Variable	Analysis level	n	Mean ± SD	Median (IQR)	Range
Posts per account	Account	685	1.87 ± 6.88	1 (1 to 1)	1-141
Current followers per account	Account	685	17,190.4 ± 89,218.0	1,098 (153 to 4,594)	0-1,294,889
Likes per post	Post	1,281	33.0 ± 111.1	3 (0 to 16)	0-2,342
Comments per post	Post	1,281	2.87 ± 15.00	0 (0 to 1)	0-377
Post age, days	Post	1,281	264.8 ± 569.7	37 (34 to 48)	33-4,242

Values are presented as mean ± standard deviation, median and interquartile range, or number and percentage, as appropriate. Account level variables were calculated using 685 unique accounts, whereas post level variables were calculated using 1,281 unique posts. Current follower counts, likes, and comments were recorded on October 15, 2025. Post age was calculated as the number of days between post publication and data extraction.

Table 2: Distribution of Account Types at the Account and Post Levels

Account type	Accounts, n (%)	Posts, n (%)
Healthcare professional	93 (13.6)	145 (11.3)
Health organization	61 (8.9)	281 (21.9)
Commercial	74 (10.8)	93 (7.3)
Influencer or advocacy	62 (9.1)	280 (21.9)
Individual	395 (57.7)	482 (37.6)
Total	685 (100.0)	1,281 (100.0)

Each account was assigned to one mutually exclusive account type, and the same classification was applied to all posts originating from that account. Percentages were calculated separately at the account and post levels.

organizations contributed 281 posts (21.9%), influencer or advocacy accounts contributed 280 posts (21.9%), healthcare professional accounts contributed 145 posts (11.3%), and commercial accounts contributed 93 posts (7.3%) (Table **2**).

Table 3: Distribution of Posts according to Caption Language

Caption language	Posts, n (%)
English	842 (65.7)
Spanish or Portuguese	159 (12.4)
Turkish	23 (1.8)
Other or uncertain	257 (20.1)
Total	1,281 (100.0)

Caption language was classified according to the predominant language of each post. Posts without sufficient caption text, posts consisting mainly of emojis or hashtags, and posts for which the predominant language could not be determined were categorized as other or uncertain.

Most posts were classified as English language content (842, 65.7%), followed by other or uncertain language content (257, 20.1%), Spanish or Portuguese content (159, 12.4%), and Turkish content (23, 1.8%) (Table **3**).

In the account clustered negative binomial generalized estimating equation model, each doubling of current follower count was associated with a 43% higher expected like count (adjusted incidence rate ratio [aIRR], 1.43; 95% CI, 1.37 to 1.50; $p < 0.001$) and a 43% higher expected comment count (aIRR, 1.43; 95% CI, 1.32 to 1.54; $p < 0.001$). Each doubling of post age was also associated with higher engagement, corresponding to a 33% higher expected like count (aIRR, 1.33; 95% CI, 1.14 to 1.56; $p < 0.001$) and a 39% higher expected comment count (aIRR, 1.39; 95% CI, 1.18 to 1.64; $p < 0.001$) (Table **4**, Figures **2** and **3**).

Using healthcare professional accounts as the reference category, health organization accounts had lower adjusted like counts (aIRR, 0.23; 95% CI, 0.11 to 0.48; $p < 0.001$) and lower adjusted comment counts (aIRR, 0.21; 95% CI, 0.06 to 0.71; $p = 0.012$). Commercial accounts, influencer or advocacy accounts, and individual accounts did not differ significantly from healthcare professional accounts in either outcome after adjustment. In particular, influencer or advocacy accounts were not significantly associated with like counts (aIRR, 0.68; 95% CI, 0.34 to 1.33; $p = 0.256$) or comment counts (aIRR, 0.50; 95% CI, 0.15 to 1.69; $p = 0.266$) (Table **4**, Figures **2** and **3**).

Compared with English language posts, Spanish or Portuguese, Turkish, and other or uncertain language posts showed no statistically significant differences in

Table 4: Account Clustered Negative Binomial Generalized Estimating Equation Models for Like and Comment Counts

Variable	Likes, adjusted IRR (95% CI)	p value	Comments, adjusted IRR (95% CI)	p value
Current follower count, per doubling	1.43 (1.37 to 1.50)	<0.001	1.43 (1.32 to 1.54)	<0.001
Post age in days, per doubling	1.33 (1.14 to 1.56)	<0.001	1.39 (1.18 to 1.64)	<0.001
Account type				
Healthcare professional	**Reference**		**Reference**	
Health organization	0.23 (0.11 to 0.48)	<0.001	0.21 (0.06 to 0.71)	0.012
Commercial	0.94 (0.47 to 1.86)	0.856	2.34 (0.38 to 14.26)	0.357
Influencer or advocacy	0.68 (0.34 to 1.33)	0.256	0.50 (0.15 to 1.69)	0.266
Individual	0.81 (0.47 to 1.39)	0.441	0.49 (0.19 to 1.23)	0.128
Caption language				
English	**Reference**		**Reference**	
Spanish or Portuguese	0.94 (0.66 to 1.33)	0.712	0.64 (0.29 to 1.40)	0.263
Turkish	0.57 (0.28 to 1.18)	0.129	0.43 (0.08 to 2.28)	0.320
Other or uncertain	0.85 (0.56 to 1.30)	0.463	1.31 (0.52 to 3.31)	0.561

IRR, incidence rate ratio; CI, confidence interval. Separate negative binomial generalized estimating equation models were constructed for like and comment counts. Account identity was specified as the clustering variable, and an exchangeable working correlation structure with robust standard errors was used. Models were mutually adjusted for current follower count, post age, account type, and caption language. Current follower count was transformed as log2(follower count + 1), and post age was transformed using log2. Therefore, the corresponding IRRs represent the expected multiplicative change in the outcome associated with a doubling of the variable. Healthcare professional accounts and English language posts were used as the reference categories.

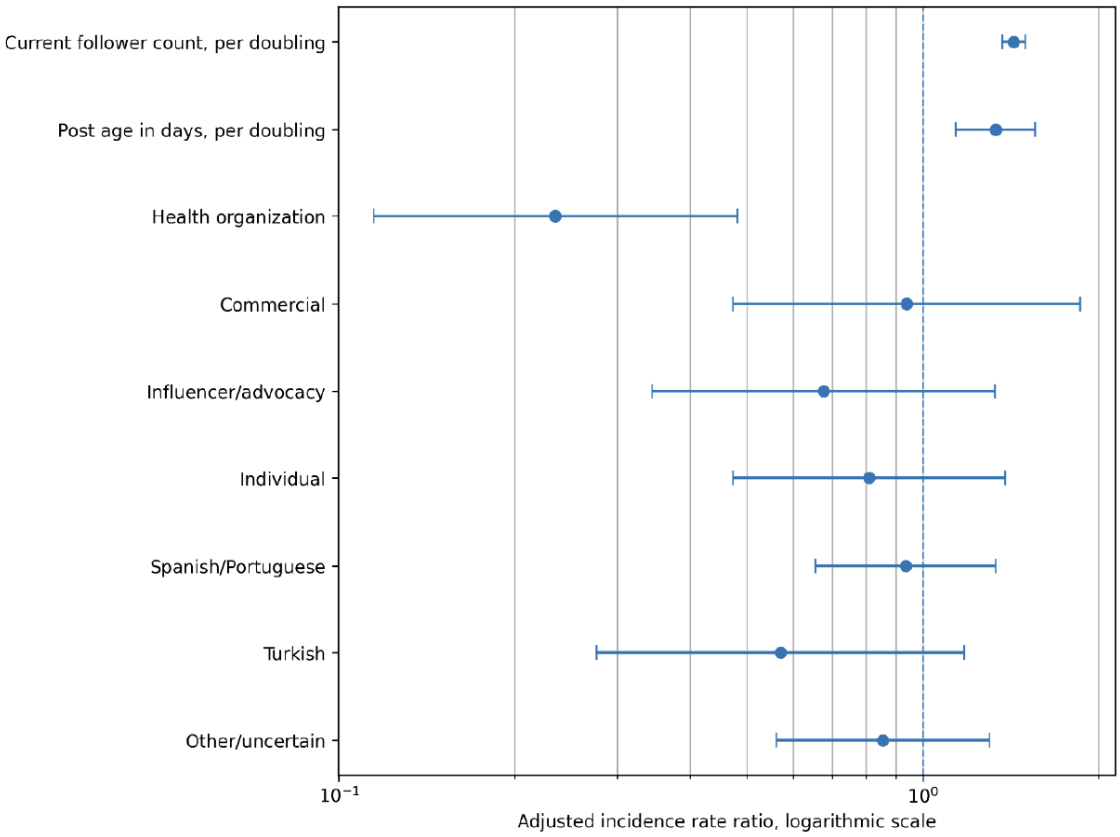

Figure 2: Adjusted associations of account and post characteristics with Instagram like counts.

Forest plot showing adjusted incidence rate ratios and 95% confidence intervals derived from the negative binomial generalized estimating equation model for like counts. Account identity was specified as the clustering variable. Models were adjusted for current follower count, post age, account type, and caption language. The vertical reference line represents an incidence rate ratio of 1.00. Healthcare professional accounts and English language posts were used as the reference categories.

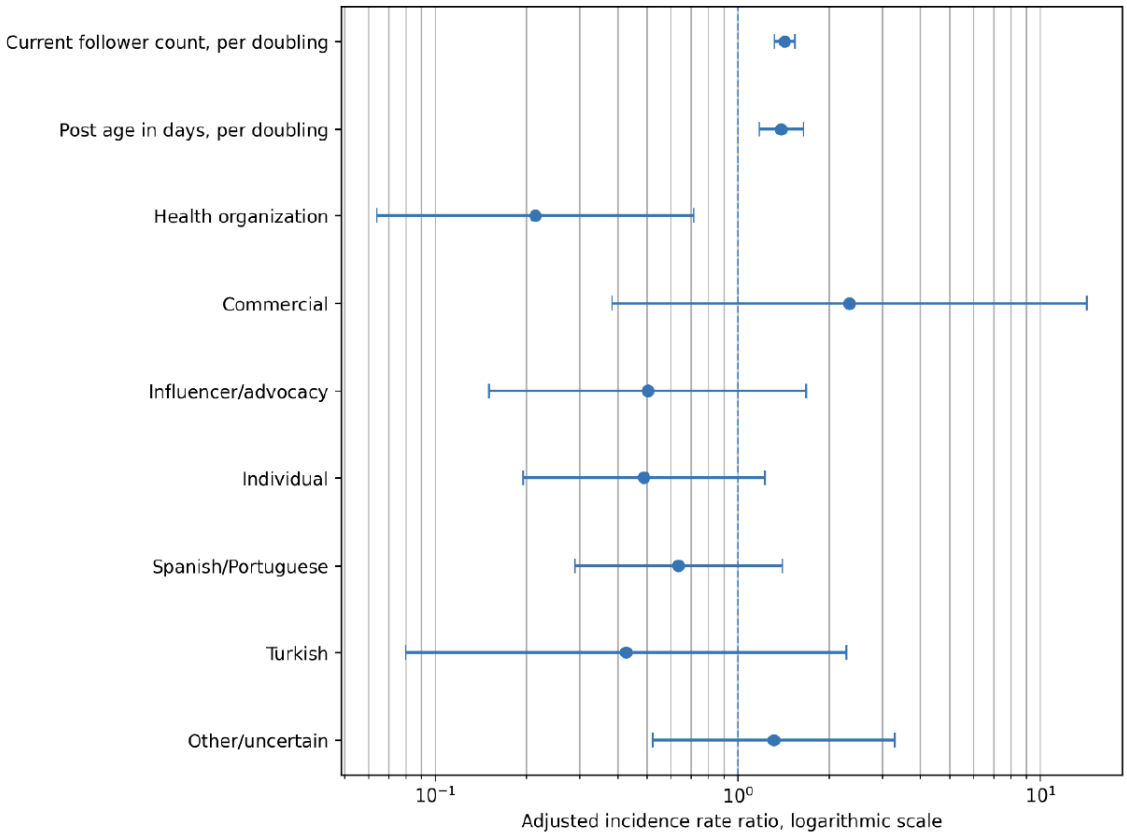

Figure 3: Adjusted associations of account and post characteristics with Instagram comment counts.

Forest plot showing adjusted incidence rate ratios and 95% confidence intervals derived from the negative binomial generalized estimating equation model for comment counts. Account identity was specified as the clustering variable. Models were adjusted for current follower count, post age, account type, and caption language. The vertical reference line represents an incidence rate ratio of 1.00. Healthcare professional accounts and English language posts were used as the reference categories.

adjusted like or comment counts (Table **4**). Negative binomial regression models using account clustered robust standard errors produced results consistent with those of the primary generalized estimating equation analyses (Supplementary Table **1**).

DISCUSSION

Our study provides a descriptive and comparative analysis of 1,281 HPV related Instagram posts retrieved through four predefined hashtags. After accounting for repeated posts from the same account and differences in post age, current follower count was positively associated with both like and comment counts. Post age was also positively associated with engagement, indicating that posts with longer exposure periods had accumulated more interactions by the extraction date. In contrast to the initial analysis, influencer or advocacy accounts did not differ significantly from healthcare professional accounts after adjustment. Caption language was not independently associated with either outcome. These findings suggest that engagement within this hashtag based sample was primarily associated with measurable account and temporal characteristics rather than with a consistent advantage for any specific nonprofessional account category.

Current follower count was positively associated with both like and comment counts. This finding is consistent with the general principle that accounts with larger audiences have a greater potential pool of users who may encounter and interact with posted content. Previous literature has emphasized that engagement with health information on social media is influenced not only by message characteristics but also by platform structure, account characteristics, and audience related factors [6,7,14]. Nevertheless, the follower counts in the present study were measured on the common data extraction date rather than at the time each post was published. They therefore cannot reconstruct the audience size, algorithmic distribution, impressions, or actual reach of an individual post at the time of dissemination. Accordingly, the observed association should be interpreted as a relationship between current account size and accumulated

engagement, rather than as evidence that follower count caused greater historical visibility.

Post age was independently associated with both like and comment counts. This finding is methodologically important because engagement metrics accumulate over time and posts published several years before data extraction have had substantially longer opportunities to receive interactions than recently published posts. Previous reviews of social media use in health communication have emphasized the dynamic nature of online content and the difficulty of interpreting engagement metrics obtained at a single cross-sectional time point [15-19]. In the present study, adjustment for the exact number of days between publication and data extraction provided a more precise temporal control than publication year alone. However, post age should not be interpreted as a measure of sustained public interest, because the timing of individual likes and comments was unavailable and Instagram engagement is typically not distributed uniformly throughout a post's lifetime. The positive association therefore indicates greater accumulated engagement among older posts rather than a continuing increase in interaction over time.

Account type was not a consistent marker of engagement after adjustment. In particular, influencer or advocacy accounts did not have significantly different like or comment counts compared with healthcare professional accounts. This finding does not support the previous interpretation that visibility outweighed professional authority. Account type was derived from publicly available profile information and represents an operational classification rather than a validated measure of expertise, credibility, scientific accuracy, or professional authority. Previous studies have described both the potential contribution of healthcare professionals to correcting misinformation and the practical challenges associated with professional participation on social media [15,20]. Other work has also shown that influencer based health communication may attract substantial attention, although engagement does not by itself establish informational quality or clinical reliability [9,21]. Accordingly, differences between account categories should be interpreted as descriptive patterns of interaction within the sampled posts, not as evidence that one account type communicated more accurately or exerted greater public health influence.

Health organization accounts had lower adjusted like and comment counts than healthcare professional accounts. This association should be interpreted cautiously because the study did not assess posting frequency outside the sampled hashtags, content format, visual design, message framing, use of video, paid promotion, audience demographics, or algorithmic distribution. Previous studies have noted that institutional health communication may be constrained by formal language, organizational approval processes, and less personalized communication styles, whereas user generated or experience based content may be perceived as more relatable [10,11,22]. However, the present dataset does not allow these mechanisms to be tested directly. The lower engagement observed for health organization accounts therefore identifies a descriptive difference within the sampled posts rather than demonstrating inferior communication quality, lower credibility, or reduced effectiveness in improving HPV related knowledge or preventive behavior.

Caption language was not independently associated with like or comment counts after adjustment. Although language may influence the accessibility, cultural relevance, and interpretation of HPV related health information, the present study did not assess the geographic location, nationality, cultural background, or health literacy of account holders or audiences. Previous research has emphasized that HPV communication and vaccine related attitudes may vary according to social, cultural, and informational contexts [18-22]. However, caption language alone cannot reliably represent these factors. The Turkish language subgroup comprised only 23 posts, resulting in imprecise estimates and limited statistical power. Therefore, the absence of a significant association should not be interpreted as evidence that language has no role in social media engagement, and no conclusions regarding regional differences can be drawn from these data.

Social media engagement should be distinguished from clinical or public health impact. Likes and comments indicate that users interacted with a post, but they do not demonstrate improved HPV knowledge, correction of misinformation, increased vaccine acceptance, higher vaccination uptake, greater participation in cervical cancer screening, or changes in preventive behavior. Previous studies have shown that social media exposure may be associated with HPV vaccine attitudes and decision making, while targeted digital campaigns may support awareness and misinformation correction [18,20-22]. Nevertheless, such outcomes require direct assessment using measures of knowledge, attitudes, intentions,

vaccination status, or screening participation. Because the present study evaluated only platform level interaction metrics, its findings should be interpreted as describing patterns of online engagement rather than demonstrating an effect on cancer prevention or population health.

The main strengths of this study are the use of a reproducible post level dataset, adjustment for post age, and statistical modeling that accounted for repeated posts from the same account. In addition, like and comment counts were analyzed as count outcomes, and the findings were supported by a sensitivity analysis.

This study also has limitations. The sample was restricted to publicly accessible posts retrieved through four predefined hashtags and may not represent all HPV related content on Instagram. Deleted, private, restricted, or inaccessible posts could not be included. Current follower count may differ from the follower count at the time of publication. Account type and caption language were manually classified, and formal interrater reliability was not calculated. Content quality, accuracy, format, audience characteristics, algorithmic exposure, paid promotion, and direct public health outcomes were not assessed. Therefore, the findings should be interpreted as associations within the retrieved sample rather than causal or platform wide effects.

CONCLUSION

In this hashtag based sample of HPV related Instagram posts, higher current follower count and greater post age were associated with higher like and comment counts. After adjustment for account level clustering and measured covariates, influencer or advocacy accounts did not differ significantly from healthcare professional accounts. These findings describe engagement patterns within the retrieved sample and should not be interpreted as evidence of content quality, professional authority, historical post visibility, or direct public health impact.

ETHICAL CONSIDERATIONS

This study was approved by the Ethics Committee of Afyonkarahisar Health Sciences University (approval no. 2025/13, dated October 3, 2025). The study involved the secondary analysis of publicly accessible Instagram posts and associated metadata and did not include direct interaction with account holders. No private or restricted accounts were accessed. Usernames, profile links, and other direct account identifiers were not reported in the manuscript, tables, or figures. Because the study did not involve recruitment, intervention, or direct participation, informed consent was not required.

AVAILABILITY OF DATA AND MATERIALS

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

CONSENT TO PARTICIPATE

Not applicable. The study used publicly available online data and did not involve direct human participation.

CONSENT TO PUBLISH

Not applicable. The study did not include any individual person's data or images requiring consent for publication.

FUNDING

The study received no external funding.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

ACKNOWLEDGEMENTS

None.

SUPPLEMENTARY TABLE

The supplementary table can be downloaded from the journal website along with the article.

REFERENCES

[1] de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. Int J Cancer 2017; 141(4): 664-670. https://doi.org/10.1002/ijc.30716

[2] Sung H, Ferlay J, Siegel RL, *et al.* Global Cancer Statistics 2020: GLOBOCAN estimates. CA Cancer J Clin 2021; 71(3): 209-249. https://doi.org/10.3322/caac.21660

[3] Drolet M, Bénard É, Pérez N, *et al.* Population-level impact and herd effects of HPV vaccination. Lancet 2019; 394(10197): 497-509. https://doi.org/10.1016/S0140-6736(19)30298-3

[4] Bruni L, Saura-Lázaro A, Montoliu A, *et al.* HPV vaccination introduction worldwide. Prev Med 2021; 144: 106399. https://doi.org/10.1016/j.ypmed.2020.106399

[5] Patel H, Jeve YB, Sherman SM, Moss EL. Knowledge of HPV and vaccine in European adolescents. Sex Transm Infect 2016; 92(6): 474-479. https://doi.org/10.1136/sextrans-2015-052341

[6] Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media in medicine and health care. J Med Internet Res 2014; 16(2): e13. https://doi.org/10.2196/jmir.2912

[7] Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. Uses, benefits, limitations of social media in health communication. J Med Internet Res 2013; 15(4): e85. https://doi.org/10.2196/jmir.1933

[8] Wang Y, McKee M, Torbica A, Stuckler D. Health misinformation on social media. Soc Sci Med 2019; 240: 112552. https://doi.org/10.1016/j.socscimed.2019.112552

[9] Shiplo SN, Gholiof M, Sarin N, Leonardi M. Endometriosis influencers on Instagram. J Minim Invasive Gynecol 2025; 32(8): 693-700. https://doi.org/10.1016/j.jmig.2025.01.018

[10] Towne J, Suliman Y, Russell KA, *et al.* Health information via Facebook for endometriosis patients. J Minim Invasive Gynecol 2021; 28(9): 1637-1642. https://doi.org/10.1016/j.jmig.2021.02.005

[11] Bautista JR, Zhang Y, Gwizdka J. Healthcare professionals' acts of correcting health misinformation on social media. Int J Med Inform 2021; 148: 104375. https://doi.org/10.1016/j.ijmedinf.2021.104375

[12] Basch CH, Hillyer GC, Tchertkov V, Basch CE. HPV on YouTube: an opportunity for education and public health. Vaccine 2017; 35(33): 4258-4262.

[13] Ortiz RR, Smith A, Coyne-Beasley T. A systematic literature review to examine the potential for social media to impact HPV vaccine uptake and awareness. Hum Vaccin Immunother 2019; 15(7-8): 1465-75 https://doi.org/10.1080/21645515.2019.1581543

[14] Kata A. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. Vaccine 2012; 30(25): 3778-89. https://doi.org/10.1016/j.vaccine.2011.11.112

[15] Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014; 39(7): 491-520.

[16] Daley EM, Perrin KM, McDermott RJ, McFarlane M, Lang D, Collins T, *et al.* The psychosocial burden of HPV infection: a systematic review of the literature. J Low Genit Tract Dis 2017; 21(2): 95-101.

[17] Kosenko KA, Hurley RJ, Harvey JA. Sources of stigma related to HPV and cervical cancer: insights from interviews with women. Women Health 2016; 56(4): 428-437.

[18] Vraga EK, Brady SS, Gansen C, Khan EM, Bennis SL, Nones MN, *et al.* A review of HPV and HBV vaccine hesitancy, intention, and uptake in the era of social media and COVID-19. eLife 2023; 12: e85743. https://doi.org/10.7554/eLife.85743

[19] Bravo CA, Walker MJ, Papadopoulos A, *et al.* Social media use in HPV, cervical cancer and cervical screening related research: a scoping review. Prev Med 2024; 179: 107798 https://doi.org/10.1016/j.ypmed.2023.107798

[20] Lin Z, Chen S, Su L, Liao Y, Chen H, Hu Z, *et al.* Influences of HPV disease perceptions, vaccine accessibility and information exposure on social media on HPV vaccination uptake among 11,678 mothers with daughters aged 9-17 years in China: a cross sectional study. BMC Med 2024; 22: 328. https://doi.org/10.1186/s12916-024-03538-1

[21] Sundstrom B, Cartmell KB, White AA, Well H, Pierce JY, Brandt HM. Correcting HPV vaccination misinformation online: evaluating the HPV Vaccination NOW social media campaign. Vaccines (Basel) 2021; 9(4): 352. https://doi.org/10.3390/vaccines9040352

[22] Liebermann E, Kornides M, Matsunaga M, Lim E, Zimet G, Glauberman G, *et al.* Use of social media and its influence on HPV vaccine hesitancy: US national online survey of mothers of adolescents, 2023. Vaccine 2025; 44: 126571 https://doi.org/10.1016/j.vaccine.2024.126571

Received on 26-05-2026 Accepted on 29-06-2026 Published on 28-07-2026

https://doi.org/10.30683/1929-2279.2026.15.20