In press at AIDS (https://journals.lww.com/aidsonline/pages/default.aspx)
doi:10.1097/QAD.0000000000002018

HIV Messaging on Twitter: An Analysis of Current Practice and Data-Driven Recommendations

Running head: HIV Messaging on Twitter

Sophie LOHMANN*

Benjamin X. WHITE

Zhen ZUO

Man-pui Sally CHAN

Alex MORALES

Bo LI

Chengxiang ZHAI

Dolores ALBARRACÍN

* Correspondence can be directed to Sophie Lohmann, Psychology Department, University of Illinois Urbana-Champaign, 603 East Daniel Street, Champaign, IL 61820. (Email: lohmann2@illinois.edu).

This project was partially supported by NIH grant NIH 1 R56 AI114501-01A1 to Dolores Albarracín. We have no conflicts of interest to disclose.

Word count: 3,499

Abstract

Objectives: Social media messages have been increasingly used in health campaigns about prevention, testing, and treatment of HIV. We identified factors leading to the retransmission of messages from expert social media accounts to create data-driven recommendations for online HIV messaging.

Design and Methods: We sampled 20,201 HIV-related tweets (posted between 2010 and 2017) from 37 HIV experts. Potential predictors of retransmission were identified based on prior literature and machine learning methods and were subsequently analyzed using multilevel negative binomial models.

Results: Fear-related language, longer messages, and including images (e.g., photos, gif, or videos) were the strongest predictors of retweet counts. These findings were similar for messages authored by HIV experts as well as messages retransmitted by experts but created by non-experts (e.g., celebrities or politicians).

Conclusions: Fear appeals affect how much HIV messages spread on Twitter, as do structural characteristics like the length of the tweet and inclusion of images. A set of five data-driven recommendations for increasing message spread is derived and discussed in the context of current CDC social media guidelines.

Keywords: HIV; health communication; health education; social media; Twitter messaging

Introduction

Posting messages on Twitter is free, fast, and holds the potential of reaching up to 330 million monthly active users [1] (20% of US Americans report using Twitter [2]) – it is no wonder that health professionals are using Twitter and other social media to disseminate messages about HIV. As early as 2010, almost a third of sexual health social media campaigns had a Twitter presence [3]. Messages featured on these accounts are endorsed by experts, which may make them more reliable than messages posted by lay users. To date, however, no study has systematically examined how much these expert-vetted messages in the area of HIV are spread online. Because sharing health content is critical to public health, messages need to be disseminated sufficiently widely for audiences to receive them in the first place. Reposting and thus re-disseminating content is a defining feature of most social media platforms including Twitter, and a marker of posting success. For example, if user B follows user A on Twitter, meaning that tweets posted by A will appear on B's Home timeline. If A posts a message that B retweets (i.e., shares), the potential audience of that particular message then equals to the total number of A's followers plus the total number of B's followers. In addition to increasing message spread, sharing a message is typically a sign that the retweeters endorse the message and believe it to be relevant or interesting to their followers. How often a message is shared is thus a good metric of message success [4] both because it indicates a positive attitude towards the message and because it widens the message's potential audience. In this paper, we analyzed dissemination of over twenty-five thousand expert-delivered HIV messages on Twitter by analyzing factors that predict greater message dissemination measured as higher retweet counts. These results fill the void in health message research on social media and generate data-driven recommendations about message design for maximal reach.

Despite the increasing usage of social media messaging, the spread of these messages varies. For instance, in the first week of November 2017, the posts of @HIVGov, the official Twitter account of hiv.gov which has 375,000 followers, received fewer than 11 retweets per message (excluding one non-HIV related tweet that was retweeted from @HealthCareGov). Similarly, 35% of Facebook posts from community-based health organizations received zero likes [5], and an analysis of charitable organizations on Facebook found that particularly crucial messages such as calls for HIV-related action were shared less than other messages [6]. Other messages and online campaigns, however, are more successful and illustrate the potential that social media messages offer. How can this potential be achieved? Prior analyses have suggested that in addition to content characteristics, surface characteristics of online messages can be influential in determining perceived credibility [7] and message dissemination [8,9]. For instance, focus groups of female Black college students have recommended including pictures and other fear-inducing visuals in media messages for HIV prevention [10]. In non-HIV specific tweets, features such as hashtags and URLs, as well as higher numbers of followers per account, were associated with higher retweet counts [11].

Retweets Versus Non-Retweets

On social media, traditional notions of message source become complex because there can be multiple layered sources [7,12]: If an influential non-expert like Elton John or Hillary Clinton posts a message, the local health department retweets it, and then somebody else retweets it from the health department, are they influenced by the reputation of the health department, by Elton John's reputation, or by the reputation of both sources? On social media, health experts therefore do not only have to choose which messages they want to post, but also which existing messages they want to retweet. Our study assesses which factors predict message retweets depending on which source the message comes from. Because source information is ambiguous when another account's message is retweeted directly, we excluded direct retweets from our analyses whenever we were interested in source factors like follower count.[1] To obtain a clear split between experts' posts and non-experts' posts, as well as between retweets and original messages, we divided the data into three groups: Original messages posted by the experts in our sample, messages they had retweeted from other HIV experts that we did not sample for, and messages they had retweeted from non-experts (incl. celebrities, politicians, and general health accounts).

In this paper, we combined theory-driven variables based on prior literature with machine learning techniques to evaluate which factors influence retweet counts of expert-generated health messages to which extent. These analyses provide empirical evidence for recommendations to increase the spread of HIV-related health messages on social media.

Methods

Data Collection

[1] Note that this applies only to unmodified retweets. In contrast, if a health expert in our sample replied to another account's message, that was a *quote tweet*. In that case, the two tweets will show up stacked on Twitter, but the reply is treated as its own separate tweet (meaning that words from the other user's original message do not appear in our data, and the reply has its own retweet count separate from the original message's retweet count).

We trained six research assistants to find HIV experts who also are Twitter users by searching conference programs, NIH staff directories, NGO websites, and HIV-related hashtags. The accounts belonged to either (a) individual experts currently working in an HIV-related area or (b) an HIV-specific organization that was either local to the U.S. (e.g., San Francisco AIDS Foundation) or operating globally (e.g., UNAids). We also reviewed the Twitter friends and followers of these accounts to find new accounts. The classifications as (non)experts were reliable (κ = .74-.79). We identified the expertise of retweets' accounts using a mix of manual coding and a supervised classifier (see Supplement). After removing duplicates, our final list of accounts included 109 individual experts and 249 expert organizations. We then selected a random sample of 20 individual and 20 expert accounts and used their Twitter usernames to get their most recent tweets (i.e., up to 3,200 tweets per username) via Tweepy, a public Python library for accessing the Twitter API. On January 24, 2017 we retrieved 69,784 tweets posted between 2010 and 2017, along with retweet and favorite counts of each tweet and each account's numbers of friends (i.e., how many accounts this user was following) and followers. All user information was publicly visible on the accounts' Twitter profiles at the time of data collection. This project used publicly available secondary data and was thus not considered human subjects research by our Institutional Review Board.

Data Filtering

To identify which tweets were HIV-related, we used Support Vector Machines (SVM), a supervised machine learning technique. Trained research assistants manually annotated 900 tweets from the selected accounts to classify whether each tweet was about

HIV and/or about other STIs. Based on these human annotations, we developed two SVM models (one to predict whether a new tweet is about HIV, one for STIs). A direct comparison between the SVM classifier's predicted values (using 10-fold cross validation) and human-annotated values showed satisfactory performance, HIV: precision (true positives divided by all tweets classified as positive) = .87, recall (true positives divided by all tweets that really are positive) = .89, accuracy = .89, and STIs: precision = .70, recall = .88, accuracy = .97. We thus used the SVM classifier's predicted values to exclude tweets that were neither about HIV nor other STIs, thus removing other topics that garner many retweets (e.g., politics) to assure the validity of our data. We excluded three accounts which posted mainly in Spanish and accounts which were unable to classify as expert or nonexpert due to account deactivation/deletion or blank description on profile. We also excluded two outliers which severely skewed the results of the analyses (one tweet with 18,719 retweets posted by @rihanna and one with 33,467 retweets posted by Barack Obama on the @POTUS account). We also excluded several tweets that were exact or near duplicates of each other (defined as Levenshtein distance > 85). For example, a user might have posted the same tweet multiple times. We aggregated these duplicates to represent them in the data only once by averaging their retweet counts. The final sample included 20,201 tweets posted between 2010 and 2017, most of which (46%) were posted in 2016 (see Figure 1).

Selection of Predictors

Based on prior literature, we hypothesized sentiment-invoking, longer, image-containing messages that utilize hashtags or URLs to achieve higher retweet counts. For emotional connotation, we used the NRC Word-Emotion Association Lexicon (EmoLex), a validated emotion-association and sentiment lexicon containing eight emotions (joy, sadness, anticipation, surprise, trust, fear, anger, disgust) and two kinds of sentiment (positive, negative) [13,14]. The scores represent counts of how many emotion-related words appear in each tweet. Each tweet's meta-data indicated whether an external link was included and whether the message included images (pictures, gifs, and/or videos).

Additional predictors included whether the tweet contained HIV-related words [15], number of emojis (counted using a byte-to-emoji dictionary [16]) whether the account belonged to an individual or an organization, whether (in the case of organizations) it had a medical or academic vs. a community focus, and a number of binary content domain variables: Teenagers and/or young adults, Black and/or Latino/a populations, transgender individuals, men who have sex with men, HIV, or other STIs (see section "Data Filtering"). Missing data for 90 tweets on two variables and missing account information for three accounts were imputed through random forest models [17,18] (see Supplement).

Variable Importance Tests

To identify the most influential variables, we used a machine learning technique called *random forests* with ten-fold cross-validation [17]. In a single decision tree, cases are split into categories at each step, based on some explanatory variable: For example, in the first step, they might be split into tweets with and without a hashtag (two groups). In the second step, these groups could be split into tweets coming from accounts with <100 vs. ≥ 100 followers (four groups), making the groups smaller and more specific with each step. For a random forest, a (typically large) number of these trees are grown, each of which makes a prediction for how many retweets a given tweet has. Then one takes the average of all these predicted values. Random forest approaches are appealing because they do not require distribution assumptions and because they can account for interactions and nonlinear relations between factors [19]. In these analyses, our sample was reduced by 102 because of missing data that the random forest method could not accommodate.

Next, we performed Gini impurity tests and permutation tests (variable importance tests [17]) to assess which variables were driving the prediction. A high value on the impurity test for variable X means that including X as a predictor makes the groups at the end of a decision tree more homogenous, meaning that the grouping is valid and suggesting that X is an important predictor. In a permutation test, the explanatory variable X is shuffled so that the values are ordered randomly. If reordering the information in the variable reduces prediction validity, the original order of the information presumably carried meaning, implying that the variable has high importance. Higher scores imply higher importance.

Negative Binomial Models

Random forest models are typically considered “black box” models because it is difficult to interpret which levels of which predictors lead to a high versus low outcome [19]. For easier interpretation and to obtain effect sizes, we then built regression models using the identified variables. Because the data were highly skewed and overdispersed count data, we used multilevel negative binomial models [20]. Each tweet was nested within the Twitter username that had posted the tweet. The coefficients from negative binomial regressions can be exponentiated to obtain incidence rate ratios (IRRs).

Results

The retweets from non-experts received much higher retweet counts than messages retweeted from HIV experts, which in turn were retweeted more often than original messages posted by our sample of HIV experts (Table 1). This latter difference is presumably a result of selection bias: By definition, a retweet is a message that has already been disseminated, so only successful messages show up in that sample. In contrast, the original messages include all tweets that were never disseminated – in fact, 54% of the original messages received zero retweets. This low retweet count was unlikely to be due to low follower counts – on average, the sampled accounts had *M* = 2,648.11 followers, *SD* = 4,509.54, *Median* = 1,025, and thus had relatively large potential audiences. For ease of interpretation, we present analyses separately for the three groups; the full model with group-by-variable interaction terms can be found in the supplement (Figure S1 and Table S2).

The impurity and permutation tests showed that follower and friend counts of the Twitter user (account-level predictors) and word count, HIV content, and emotional sentiment (tweet-level predictors) emerged as the most influential predictors of retweet counts (see supplement, Table S1). Anger, fear, trust, and general positive sentiment all appeared as important sentiment predictors. Due to multiple meanings of a single word, multicollinearity was present for several of our emotion predictors that caused counterintuitive suppression effects (e.g., $r_{\text{fear-anger}} = .64$, $r_{\text{positive-trust}} = .53$). Therefore, we included only one positive and one negative predictor in our models. We used fear rather than anger due to a robust literature on fear appeals for persuasion in health domains and because in the context of our tweets, the words that showed up in both dictionaries (e.g., “epidemic”, “stigma”, “disease”) seemed more appropriately described as fear-related than anger-related. We chose trust instead of positive sentiment as it may increase individual response efficacy and thus better predict persuasion. In addition, the trust dictionary words may have communicated enhanced reputation of the message (e.g., “fact”, “center”, “important”).

Incidence Rate Ratios

Based on these results and prior literature on tweet characteristics, we created a model with fear, trust, word count, hashtag count, visual content, URL use, and HIV content as tweet-level predictors. Tweets with more fear-related terms, longer tweets, and tweets that included a picture, gif, or video were associated with higher retweet counts (Table 1). Specifically, for each fear-related word, the model predicted a 5% increase in retweets. Each additional 10 words were associated with an estimated increase of 18% in retweets, and visual content was associated with over 80% more retweets. The effect of including a URL was negative and was significant only among the expert retweets. All these effects were also significant (and usually even stronger) in the smaller sample of non-expert retweets. For expert retweets, the effect of trust words was negative and the effect of hashtag was positive. Then, we fit a second model with follower and friend counts as account-level predictors for the original messages. Neither follower count, *IRR* = 1.00, 95% CI [1.00, 1.00], nor friend count, *IRR* = 1.00, 95% CI [1.00, 1.00], were significant predictors.

Discussion

The results revealed that fear appeals, longer tweets, and visual content predicted higher retweet counts, whereas including a URL was actually associated with fewer retweets among expert retweets (non-significantly so among original messages). Confirming our hypotheses, the variable importance analyses indicated that tweet length predicted retweet counts. The number of followers and number of friends were important predictors, although their effects disappeared in the regression analyses after nesting the data within accounts.

Negative (fear-related), but not positive (trust-related) language was associated with more retweets, which mirrors prior literature on social media dissemination [21–23] as well as a wider literature on fear appeals that suggests these messages may be more persuasive [24]. Messages that are completely neutral in tone do not seem well-suited to attaining very high numbers of retweets. Therefore, fear appeal messages may be useful for spreading content on Twitter. Positive language may have failed to increase retweets due to the audience: negative messages may be more persuasive on average, while positive messages may be more persuasive to only some groups [25]. Considering Twitter demographics, the proportion of people who see a message may be more likely to fall into the at-risk rather than clinical population.

Second, we found that longer messages garner more retweets, at least within a 140-character limit. This finding is inconsistent with the CDC guidelines on social media use, which recommend to "keep messages short… Use fewer characters than allowed to make sharing easy" [26]. Ironically, it seems that using too few words discourages rather than encourages sharing of health messages. Shorter messages may not provide enough detail for readers to determine content credibility, as previous work has shown that simply being from a reputable source does not offset credibility-harming effects in the message itself [27].

Third, including visual content was strongly associated with increased dissemination in our analyses and may be attributed to increased engagement with message content or as a method of providing more information than 140 characters allow, thus mirroring findings regarding tweet length. Fourth, the effect of hashtags was significant only among the expert retweets but overall indicated that more hashtags may be related to more dissemination, which may be attributed to increased viewership in communities interested in HIV content.

Fifth, messages that linked to an external website occasionally received fewer retweets, especially when sharing another expert's message. This may seem surprising until one considers that due to the space limitations on Twitter, many tweets with links function as "teasers" rather than full-fledged messages. If the tweet merely mentions that there are "5 new ways to prevent HIV" and users are required to open a link to learn about the actual methods, many users may ignore the message instead of going through the effort of opening and reading the linked article. Alternatively, experts may use links especially for content that may seem uninteresting to a wider audience, like academic articles. In concordance with the CDC guidelines, we recommend ensuring that the message itself provides interesting information, without relying on linked pages to deliver the central content.

Finally, even the most-disseminated messages from HIV experts reached nowhere as many retweets as messages from Obama, Rihanna, Bill Gates, Elton John, or the WHO did. The two most popular tweets in our sample were both from non-experts and had to be excluded because they heavily skewed results. For particularly important messages that need to reach a large audience, collaborations with other accounts that already have a large social media following are thus worth considering.

Limitations

Our results need to be interpreted in light of our methods and their limitations. The data were non-experimental and cross-sectional, therefore the results cannot offer causal conclusions. It is possible that the relations we find are produced by unexplained third variables, although prior research with similar findings may add credence to our results. Additionally, this analysis focused on Twitter, and whether these results transfer to other social media platforms is an empirical question that future research needs to address. One particular limitation arises from Twitter's recent expansion of the character limit from 140 to 280. Our data was collected prior to this expansion, and although our results suggest that longer tweets fare better, it is unclear whether that result also applies beyond 140 characters. Nonetheless, variations in character length are unlikely to invalidate the effects of URL, visual content, or emotional language on message spread.

Finally, health experts use social media for a variety of purposes [28], and retweets may not be a suitable success metric for all purposes. For instance, if a message is intended to reach a few specific individuals or build connections with another expert, a high retweet rate will not be necessary. Nevertheless, most messages in our database appeared geared towards a wider audience, making retweets a valid outcome metric for these tweets.

Conclusions

In order for a health message to affect knowledge or behavior, it first needs to be received. Our analysis focused on this first step by examining retweet counts, where more retweets spread the message to a wider audience, thus leading to a higher probability of being received. Future studies should examine if the same variables that contribute to message dissemination also influence knowledge or behavior change in the audience. Also, as attitudes, knowledge, and behavior are difficult to measure on a per-message level on social media, laboratory studies and online experiments may complement social media analyses. Later studies can also contribute to analyzing the intention of each tweet (e.g., aimed at increasing knowledge versus changing behavior), analyzing the content of the pictures that increase retweets, and expanding the analyses to include other languages.

We propose a set of recommendations that HIV experts and organizations may use to shape their social media presence (Table 2). When experts post their own messages and follow recommendations 1-4, our model would suggest that these steps lead to up to 3-fold increases in expected retweets (when using three fear-related terms, using 30 instead of 10 words, including a picture, and not requiring a link). By following these recommendations, their health messages may reach a wider audience and thus potentially be more effective at changing knowledge or behavior.

Acknowledgments and Disclosures

We are grateful to our research assistants for help in identifying expert accounts (Berkelee A. Asare, Yujing Chen, Colin M. Harmony, Ziyi Lai, Janina Rojas, Kenneth Sanchez, Shiying Sun, Ailing Wang), help in manually annotating tweet content domains (Yujing Chen, Kenneth Sanchez, Ailing Wang), and help with a preliminary literature review (Nupoor Gandhi). This project was partially supported by NIH grant NIH 1 R56 AI114501-01A1 to Dolores Albarracín. We have no conflicts of interest to disclose.

References

1 Statista. **Number of mobile monthly active Facebook users worldwide from 1st quarter 2009 to 1st quarter 2016 (in millions)**. Statista. 2016.http://www.statista.com/statistics/277958/number-of-mobile-active-facebook-users-worldwide/ (accessed 12 Aug2017).

2 Duggan M. **Mobile messaging and social media 2015**. ; 2015. http://www.pewinternet.org/2015/08/19/mobile-messaging-and-social-media-2015/

3 Gold J, Pedrana AE, Sacks-Davis R, Hellard ME, Chang S, Howard S, *et al.* **A systematic examination of the use of online social networking sites for sexual health promotion**. *BMC Public Health* 2011; **11**:583.

4 Centers for Disease Control and Prevention. **Social media guidelines and best practices: CDC Twitter profiles**. ; 2011. https://www.cdc.gov/socialmedia/tools/guidelines/pdf/twitterguidelines.pdf

5 Ramanadhan S, Mendez SR, Rao M, Viswanath K. **Social media use by community-based organizations conducting health promotion: A content analysis**. *BMC Public Health* 2013; **13**:1129.

6 Huang YC, Lin YP, Saxton GD. **Give me a like: How HIV/AIDS nonprofit organizations can engage their audience on Facebook**. *AIDS Educ Prev* 2016; **28**:539–556.

7 Sundar SS. **The MAIN model: A heuristic approach to understanding technology effects on credibility**. In: *Digital media, youth, and credibility*. Metzger MJ, Flanagin AJ (editors). . Cambridge, MA: MIT Press; 2008. pp. 73–100.

8 Naveed N, Gottron T, Kunegis J, Alhadi AC. **Bad news travel fast**. In: *Proceedings of the 3rd International Web Science Conference on - WebSci '11*.New York, New York, USA: ACM Press; 2011. pp. 1–7.

9 Jenders M, Kasneci G, Naumann F. **Analyzing and predicting viral tweets**. In: *Proceedings of the 22nd International Conference on World Wide Web - WWW '13 Companion*.New York, New York, USA: ACM Press; 2013. pp. 657–664.

10 Chandler-Coley R, Ross H, Ozoya O, Lescano C, Flannigan T. **Exploring Black college females' perceptions regarding HIV prevention message content**. *J Health Commun* 2017; **22**:102–110.

11 Suh B, Hong L, Pirolli P, Chi EH. **Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network**. *Proc - Soc 2010 2nd IEEE Int Conf Soc Comput PASSAT 2010 2nd IEEE Int Conf Privacy, Secur Risk Trust* 2010; :177–184.

12 Lee JY, Sundar SS. **To tweet or to retweet? That is the question for health professionals on Twitter**. *Health Commun* 2013; **28**:509–524.

13 Mohammad SM, Turney PD. **Crowdsourcing a word-emotion association lexicon**. In: *Computational Intelligence*.; 2013. pp. 436–465.

14 Mohammad SM, Turney PD. **Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon**. *CAAGET '10 Proc NAACL HLT 2010 Work Comput Approaches to Anal Gener Emot Text* 2010; :26–34.

15 Lohmann S, Albarracín D. **HIV dictionary [unpublished data file]**. 2018.

16 Azhar H. **Prismoji emoji dictionary [data file]**. 2017.https://github.com/PRISMOJI/emojis/tree/master/2017.0206 emoji data science tutorial

17 Breiman L. **Random forests**. *Mach Learn* 2001; **45**:5–32.

18 Stekhoven DJ, Bühlmann P. **MissForest—non-parametric missing value imputation for mixed-type data**. *Bioinformatics* 2012; **28**:112–118.

19 Louppe G. *Understanding random forests: From theory to practice [Doctoral dissertation]*. 2014. doi:10.13140/2.1.1570.5928

20 Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, *et al.* **Modeling zero-inflated count data with glmmTMB**. ; 2017. doi:10.1101/132753

21 McLaughlin ML, Hou J, Meng J, Hu C-W, An Z, Park M, *et al.* **Propagation of information about preexposure prophylaxis (PrEP) for HIV prevention through Twitter**. *Health Commun* 2016; **31**:998–1007.

22 Brady WJ, Wills JA, Jost JT, Tucker JA, Van Bavel JJ. **Emotion shapes the diffusion of moralized content in social networks**. *Proc Natl Acad Sci* 2017; **114**:7313–7318.

23 Stieglitz S, Dang-Xuan L. **Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior**. *J Manag Inf Syst* 2013; **29**:217–248.

24 Tannenbaum MB, Hepler J, Zimmerman RS, Saul L, Jacobs S, Wilson K, *et al.* **Appealing to fear: A meta-analysis of fear appeal effectiveness and theories**. *Psychiatr Bull* 2015; **141**:1178–1204.

25 Block LG, Keller PA. **When to accentuate the negative: The effects of perceived efficacy and message framing on intentions to perform a health-related behavior**. *Source J Mark Res* 1995; **32**:192–203.

26 Centers for Disease Control and Prevention. **CDC's guide to writing for social media**. ; 2012. https://www.cdc.gov/socialmedia/tools/guidelines/pdf/GuidetoWritingforSocialMedia.pdf

27 Mitra T, Wright GP, Gilbert E. **A parsimonious language model of social media credibility across disparate events**. doi:10.1145/2998181.2998351

28 Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. **A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication**. *J Med Internet Res* 2013; **15**:e85.

Table 1

Multilevel negative binomial models of standardized effects regressing retweet count on tweet-level variables by three groups

	Range	Original messages		Retweets from experts		Retweets from non-experts	
N		13,471		5,374		1,356	
M (*SD*) retweet count		1.18 (2.93)		9.58 (68.06)		62.05 (324.05)	
		IRR	**95% CI**	**IRR**	**95% CI**	**IRR**	**95% CI**
Intercept		0.51*	[0.34, 0.76]	1.82*	[1.40, 2.36]	3.65*	[1.79, 7.44]
Fear	[0 – 5]	1.05*	[1.00, 1.10]	1.06*	[1.02, 1.10]	1.21*	[1.07, 1.36]
Trust	[0 – 5]	1.01	[0.97, 1.06]	0.96*	[0.92, 0.99]	1.03	[0.92, 1.15]
Word Count	[1 – 33]	1.02*	[1.01, 1.02]	1.04*	[1.03, 1.05]	1.04*	[1.02, 1.06]
Hashtag Count	[0 – 10]	1.02	[1.00, 1.05]	1.04*	[1.02, 1.06]	1.05	[0.98, 1.11]
Visual Content	Binary	1.87*	[1.74, 2.01]	1.35*	[1.24, 1.48]	1.87*	[1.52, 2.30]
Used URL	Binary	0.96	[0.89, 1.03]	0.91*	[0.87, 0.96]	1.12	[0.95, 1.30]
Content: HIV	Binary	1.12	[0.94, 1.34]	1.11	[0.92, 1.34]	0.67	[0.4, 1.13]

* $p < .05$

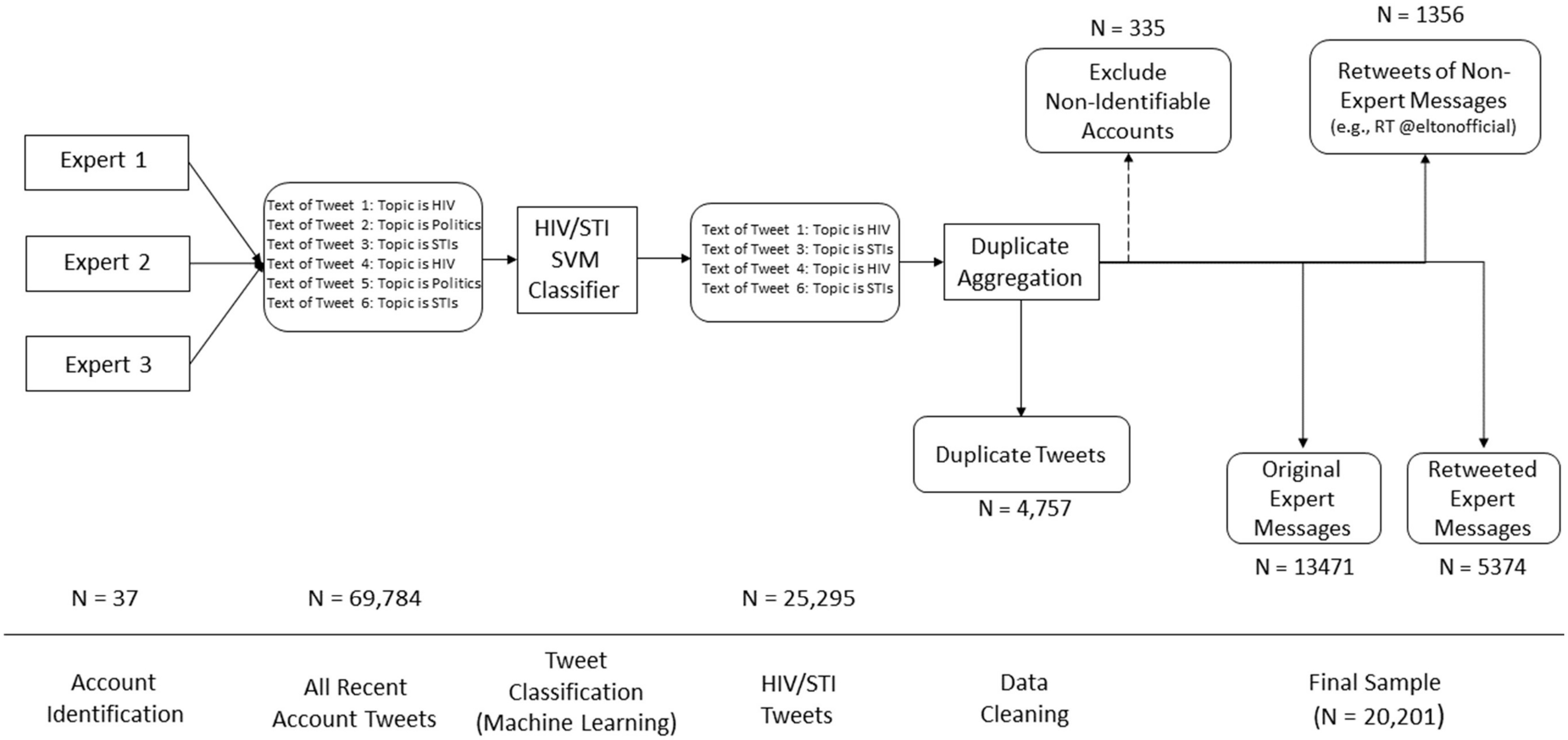
Note. Split by three groups: Original health expert tweets (nested within username), and HIV/STI-related tweets retweeted from HIV experts versus nonexperts (nested within the username that originally posted the tweet). Sample of 20,201 HIV-related tweets (posted between 2010 and 2017) from 37 HIV experts. For continuous predictors, the IRRs are scaled by one-unit increases, for instance, IRR = 1.05 means a 5% retweet increase for each additional fear-related word.

Table 2

Social Media Recommendations for Health Messages

Social Media Recommendations for Health Messages
1. Include fear appeals where appropriate.
2. Avoid short messages.
3. Include a picture, gif, or video.
4. When a message includes a link to an external page, ensure that the content of the message is still interesting and informative on its own.
5. Where possible, collaborate with existing influencers on the specific social media platform.

Figure 1. Flowchart of Process for Obtaining Tweet Sample


N = 335
Exclude Non-Identifiable Accounts
N = 1356
Retweets of Non-Expert Messages (e.g., RT @eltonofficial)
Expert 1
Expert 2
Expert 3
Text of Tweet 1: Topic is HIV
Text of Tweet 2: Topic is Politics
Text of Tweet 3: Topic is STIs
Text of Tweet 4: Topic is HIV
Text of Tweet 5: Topic is Politics
Text of Tweet 6: Topic is STIs
HIV/STI SVM Classifier
Text of Tweet 1: Topic is HIV
Text of Tweet 3: Topic is STIs
Text of Tweet 4: Topic is HIV
Text of Tweet 6: Topic is STIs
Duplicate Aggregation
Duplicate Tweets
N = 4,757
Original Expert Messages
N = 13471
Retweeted Expert Messages
N = 5374
N = 37
N = 69,784
N = 25,295
Account Identification
All Recent Account Tweets
Tweet Classification (Machine Learning)
HIV/STI Tweets
Data Cleaning
Final Sample (N = 20,201)

Supplemental Digital Content

Supplement. Text Document with Additional Methodological Details.

Table S1. Results of the Variable Importance Tests.

Table S2. Multilevel Negative Binomial Model of Standardized Effects Regressing Retweet Count on Tweet-Level Variables.

Figure S1. Marginal effects from negative binomial regression (see Table S2) for continuous predictors. 95% confidence intervals are shaded in gray.

Supplement

Topic Classifier

Nine hundred tweets were used to develop a Support Vector Machine (SVM) classification model. Using the term-frequencies normalized by the inverse document frequency (tf-idf) as the feature set, SVM transforms the feature space with a kernel which allows for better classification. Compared to other models such as logistic regression, SVMs are computationally less efficient in training because they have many more parameters [1]. Therefore, they require extensive computational power for large-scale training tasks and a large amount of training data [2]. However, SVM has been extensively used in text classification problems, and importantly, SVM resulted in better recall and precision than artificial neural network in classifying short text documents (similar to the data corpus included in the present study) [3]. Further, SVM typically retains high efficiency when applying the classification models to new testing examples.

We used a 10-fold cross-validation method to avoid overfitting, meaning that we trained the models on 9/10 of the data and tested the performance on the remaining one-tenth of the data, and repeated for each tenth of the data. We used the Python package Scikit-learn's implementation of SVM, with the default Radial Basis Function (RBF) kernel to learn the classifier, i.e., giving a label of "Yes" or "No" to each tweet, indicating whether it is about HIV or STI. The comparison test between the predicted labels and human-annotated values showed satisfactory performance of the classification models (HIV: precision = .87, recall = .89, accuracy = .89 and STI: precision = .70, recall = .88, accuracy = .97). Afterwards, we applied the classification models to determine whether each tweet was about HIV/STI and limited the dataset from 68,638 to 25,895.

Missing Data

The tweets were originally retrieved January 2017, and data on whether the message included an external link or visual content was retrieved October 2017. In the meantime, 90 tweets had become no longer inaccessible (e.g., because the account or the tweet had been deleted), therefore there were missing values on these two variables. We used random forest models to impute these 180 values [4,5]. Follower and friend counts were missing for three accounts, and we imputed their January counts based on the October counts on these variables, assuming that their followers and friends had grown at the same rate as those of the other accounts.

Variable Importance Analyses

Random forest models [4] are an ensemble learning method used mainly for classification and regression purposes. In a single decision tree, cases are split into two categories at each step, based on some explanatory variable: For example, in the first step, they might be split into tweets with and without a hashtag (two groups), in the second step, these groups could be split into tweets coming from accounts with <100 vs. ≥ 100 followers (four groups). Random forests operate by constructing a large number of these binary decision trees that grow in randomly selected subspaces of the data, and then outputting the most popular class for classification or mean prediction for regression of the individual trees. Bootstrap aggregating is used to improve the stability and accuracy of a collection of tree-structured classifiers. Ten-fold cross validation is used to find the optimal number of trees that should be grown, the number of predictors used to determine the split at each node, and the minimum size of terminal nodes (output groups) [6]. The random forest approach has several advantages with respect to other models such as classification or multivariate regression. First, it does not require distribution assumptions for explanatory variables; second, it allows for the mixed use of categorical and numerical factors; and third, it is capable of accounting for interactions and nonlinear relationships between factors [7].

Breiman [4,8] proposed to evaluate the importance of a variable $\boldsymbol{Xm}$ for predicting **Y** by adding up the weighted impurity decreases $p(t)_i(s_t,t)$ for all nodes t where Xm is used, averaged over all N_T trees in the forest:

$$Imp(X_m) = \frac{1}{N_T}\sum_{\mathrm{T}} \sum_{\mathrm{t \in T : v(s_t) = X_m}} p(t)\Delta(s_t,t)$$

$$\Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R)$$

where t_L is the left children and t_R is the right children. $pL = \frac{NtL}{Nt}$ and $pR = \frac{NtR}{Nt}$. N_T is the total sample size, and N_t is the number of samples in the parent node. Here, $i(t)$ is the Gini index. For Binary Target variable,

$$i(t) = 1 - \sum_{i=1,2} p_i^2$$

$p(t)$ is the proportion $\frac{N_t}{N}$ of samples reaching t and $v(s_t)$ is the variable used in split s_t. Features which produce large values for this score are ranked as more important than features which produce small values.

For permutation tests, according to Breiman [4,8], first, fit an initial random forest to the data. During the fitting process the out-of-bag error for each data point is recorded and averaged over the forest. Next, the values of the $\boldsymbol{X}_m$ are permuted and the out-of-bag error is again computed on this perturbed data set. The importance score for the $\boldsymbol{X}_m$ feature is computed by averaging the difference in out-of-bag error before and after the permutation over all trees. The score is normalized by the standard deviation of these differences. Features which produce large values for this score are ranked as more important than features which produce small values. Both impurity and permutation tests were applied with subsampling without replacement to correct the bias in random forest variable importance measures [9].

User Type Classifier

As explained in the main manuscript, there were posts in our sample that were not originally posted by the health experts we had sampled, but rather retweeted from someone else's account. Some were retweeted from non-HIV experts, such as celebrities or politicians. Others, however, were retweeted from other HIV experts (e.g., colleagues), obfuscating the comparison between original expert posts and non-expert retweeted posts we desired. Additionally, because we did not sample for them, these expert retweets were biased, meaning that only highly-popular tweets were included but all the tweets with few (e.g., zero) retweets were not.

Therefore, we excluded expert-authored retweets using a classifier: Based on manual codings of expert and non-expert accounts (see section Methods: Data Collection in the main manuscript), we developed a classifier with which we could sort new accounts into those same categories. We randomly collect personal descriptions from 311 Twitter users who posted a message using HIV-related words and combined it with descriptions of individual and organization experts to identify a Support Vector Machine (SVM) classification model. About fifty-two percentages of all selected Twitter users included a personal description of their profile, and we finally included 335 personal descriptions in the model development. Likewise, we used a 10-fold cross-validation method to avoid overfitting, and a user category classifier, i.e., a label of "individual expert", "organization expert", or "non-expert" is given to each

Twitter user, indicating whether a particular user is considered to be an expert in HIV based on personal description provided on his/her Twitter profile. We assessed the model performance by comparing the predicted label and the actual user category and there was a satisfactory performance of the classification of user category, i.e., precision = .80, recall = .85, accuracy = .83.

References

1 Pawar PY, Gawande SH. **A comparative study on different types of approaches to text categorization**. *Int J Mach Learn Comput* 2012; :423–426.

2 Devika MD, Sunitha C, Ganesh A. **Sentiment analysis: A comparative study on different approaches**. *Procedia Comput Sci* 2016; **87**:44–49.

3 Basu A, Walters C, Shepherd M. **Support vector machines for text categorization**. In: *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*.IEEE; 2003. p. 7 pp.

4 Breiman L. **Random forests**. *Mach Learn* 2001; **45**:5–32.

5 Stekhoven DJ, Bühlmann P. **MissForest—non-parametric missing value imputation for mixed-type data**. *Bioinformatics* 2012; **28**:112–118.

6 Liaw A, Wiener M. **Classification and regression by randomForest**. *R News* 2002; **2**:18–22.

7 Louppe G. *Understanding random forests: From theory to practice [Doctoral dissertation]*. 2014. doi:10.13140/2.1.1570.5928

8 Breiman L. **Manual on setting up, using, and understanding random forests v3.1**. ; 2002. https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf

9 Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. **Bias in random forest variable importance measures: Illustrations, sources and a solution.** *BMC Bioinformatics* 2007; **8**:25.

Table S1. Results of the Variable Importance Tests.

Variable Type	Variable	Impurity Reduction	Permutation Validity Reduction
Account	Username	693851.84	622.73
Account	Friend count	379925.82	380.98
Structural	Word count	365707.22	13.16
Account	Follower count	327406.73	691.83
Structural	Favorite count	84889.82	15.27
Content	HIV	81888.22	20.20
Structural	Hashtag count	67752.04	6.77
Structural	External URL	63787.82	13.41
Sentiment	Fear	53040.18	7.17
Account	Account type*	52781.88	106.53
Structural	Picture/gif/video	45526.74	7.02
Sentiment	Positive	33647.77	10.44
Content	HIV (dictionary-based)	32281.82	1.65
Sentiment	Anticipation	28136.94	2.19
Sentiment	Trust	23261.88	2.55
Content	Other STIs	21381.28	4.08
Sentiment	Negative	17595.35	6.07
Sentiment	Disgust	17061.72	0.18
Sentiment	Anger	7567.53	0.77
Sentiment	Sadness	6531.93	0.34
Content	Black or Latino/Latina people	5663.79	0.81
Sentiment	Joy	5199.98	0.63
Sentiment	Surprise	2816.87	0.03
Structural	Emoji count	1853.42	0.18
Content	Transgender people	1287.68	0.03
Content	Men who have sex with men	61.42	0.01
Content	Young people	45.40	0.01

* 0 = expert individual, 1 = nonmedical expert organization, 2 = expert organization with a more medical focus

Note. Higher values indicate higher importance. Sample of 20,201 HIV-related tweets (posted between 2010 and 2017) from 37 HIV experts.

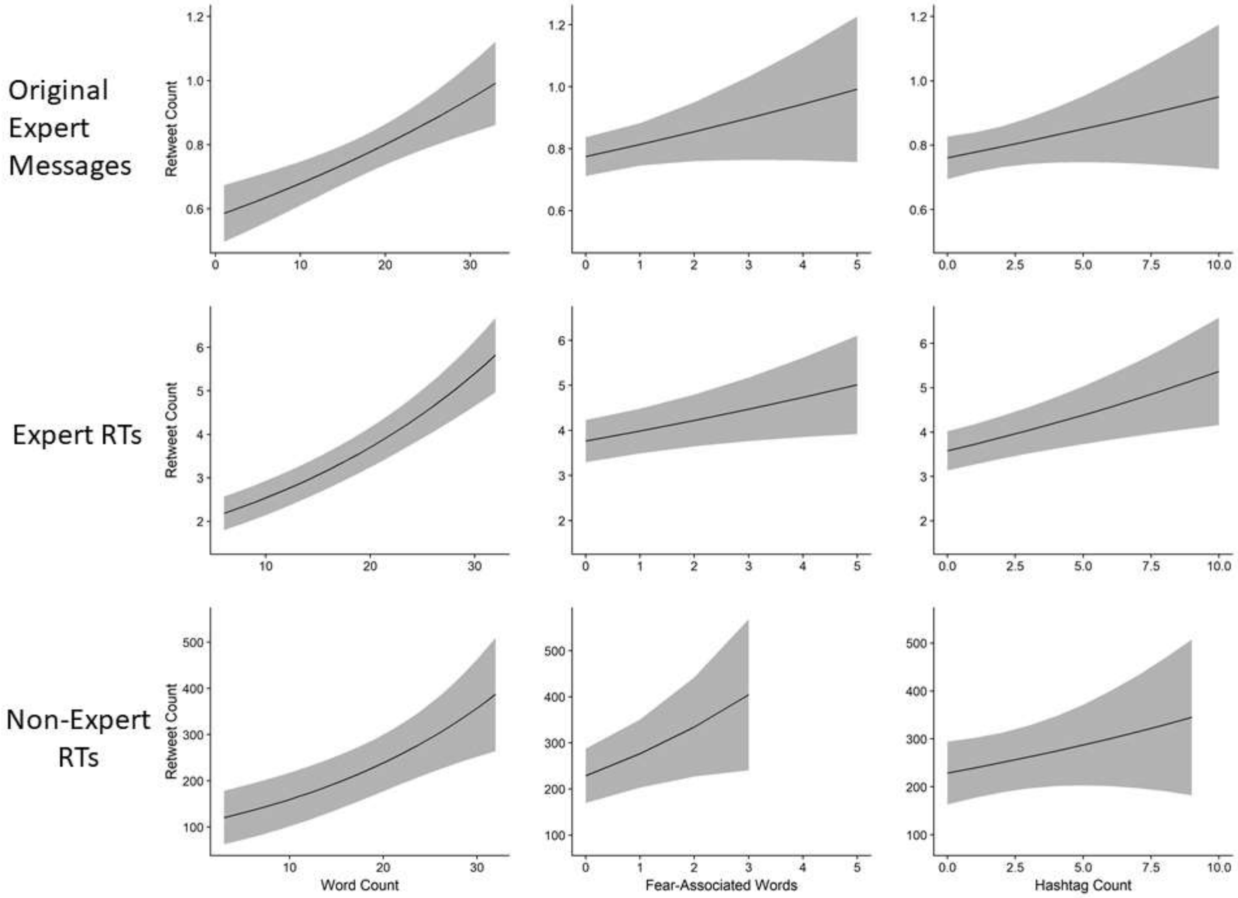
Table S2

Overall multilevel negative binomial model of standardized effects regressing retweet count on tweet-level variables

	Main effect (reference group: original tweets)		Interaction coefficient with expert retweets		Interaction coefficient with nonexpert retweets	
Parameter	IRR	95% CI	IRR	95% CI	IRR	95% CI
Intercept	0.67*	[0.52, 0.86]				
Expert retweet			2.73*	[1.84, 4.07]		
Nonexpert retweet					4.61*	[2.18, 9.76]
Fear	1.05*	[1.01, 1.10]	1.00	[0.94, 1.07]	1.17*	[1.03, 1.33]
Trust	1.01	[0.98, 1.05]	0.94	[0.88, 1.00]	1.00	[0.89, 1.12]
Word Count	1.02*	[1.01, 1.02]	1.02*	[1.01, 1.03]	1.04*	[1.02, 1.06]
Hashtag Count	1.02	[1.00, 1.04]	1.02	[0.99, 1.06]	0.98	[0.92, 1.05]
Visual Content	1.87*	[1.76, 1.99]	0.72*	[0.64, 0.81]	1.12	[0.90, 1.40]
Used URL	0.96	[0.91, 1.02]	0.94	[0.86, 1.03]	1.20*	[1.02, 1.42]
Content: HIV	1.14	[0.97, 1.33]	0.98	[0.74, 1.29]	0.61	[0.35, 1.06]

* *p* < .05

Figure S1. Marginal effects from negative binomial regression (see Table S2) for continuous predictors. 95% confidence intervals are shaded in gray.


Original Expert Messages
Expert RTs
Non-Expert RTs
Retweet Count
Word Count
Fear-Associated Words
Hashtag Count