PNAS Nexus, 2025, **4**, pgaf171
https://doi.org/10.1093/pnasnexus/pgaf171
Advance access publication 17 June 2025
Research Report

Living health-promotion campaigns for communities in the United States: Decentralized content extraction and sharing through AI

Man-pui Sally Chan[a,b], Haesung Jung[a,b], Alex Morales[c], Angela Zhang[d], Devlin O'Keefe[a,b], Sarah Joseph[e,†], Anthony Hron[e,†], Janet Davis[e,†], Tito Terry[e,†], Tiffany Peterson[e,†], Corey Herrman[e,†], Melissa Phillips[e,†], Jennifer Osborne[e,†], Kelley G. McBride[e,†], Martin Hensley[e,†], Adriana Todorov[e,†], Alain Morrissette[e,†], Georgett Watson[e,†], Ethan Knox[e,†], Erin Lark[e,†], Elisa Long[e,†], Carolina Guerrero-Lara[e,†], Timothy Rissel[e,†], Michele Raymond[e,†], Patrick Sullivan[f], Sophie Lohmann[g], Aashna Sunderrajan[g], Marta R. Durantini[a,b], Travis Sanchez[f], Chengxiang Zhai[c] and Dolores Albarracin[a,b,d,h,i,*]

[a]Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA 19104, USA
[b]Annenberg Public Policy Center, University of Pennsylvania, Philadelphia, PA 19104, USA
[c]Department of Computer Science, University of Illinois, Champaign, IL 61801, USA
[d]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA
[e]Public Health Agencies and Community-Based Organizations in the United States
[f]Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA
[g]Department of Psychology, University of Illinois, Champaign, IL 61820, USA
[h]Department of Family and Community Health, School of Nursing, University of Pennsylvania, Philadelphia, PA 19104, USA
[i]Wharton Health Care Management, Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA
*To whom correspondence should be addressed: Email: dalba@upenn.edu
†The full list of the authors and affiliations at the end of the manuscript.
Edited By John Bellettiere

Abstract

Even though health-promotion campaigns can elicit behavioral change among constituents, these initiatives are generally implemented through expensive, centralized, unsystematic, and time-consuming efforts led by creatives and officials in federal and national agencies. Can advancements in AI provide systematic methods that generate living health campaigns out of social media posts generated by communities? Here, we report the success of an innovative method to automatically select actionable HIV prevention and testing messages from decentralized content on social media (e.g. X [formerly Twitter]). The method was assessed through computational methods, an online experiment with men who have sex with men, and a field experiment involving public health agencies and community-based organizations with jurisdiction in 42 counties in the United States. The computational analyses showed that the method is computationally successful. The results of the two experiments indicated that the resulting messages are perceived as more actionable, personally relevant, and effective, and the messages are six times as likely to be posted by agencies in United States counties.

Keywords: human immunodeficiency virus (HIV), health-promotion campaigns, actionability, social media, health communication

Significance Statement

Three studies, including a series of computational analyses, an online experiment with men who have sex with men, and a field experiment with public health agencies, demonstrated that the AI method is computationally successful, and the resulting messages are perceived as more actionable, personally relevant, and effective and are six times as likely to be adopted by agencies in US counties.

Introduction

The United States has embarked on an initiative to eradicate infection with the HIV by 2030 (1) and aims to control respiratory infections such as COVID-19 and influenza. However, generating media campaigns to support these efforts remains an art often left to the intuition of advertising creatives and public health officials. Messages are crafted without a clear sense of whether health departments and communities will accept them, and their effectiveness is rarely, if ever, ascertained. A recent meta-analysis conducted by Athey et al. (2) reported that the digital campaigns costed about US$105,183 (SD = 16,872) on average. To improve on this traditional approach, this article reports the design and

Competing Interest. The authors declare no competing interests.
Received: October 14, 2024. **Accepted:** April 29, 2025

validation of a living method using decentralized, community-based content to systematically generate and scale up public health messages in counties in the United States. The method, which is based on AI, was assessed computationally and then through an online experiment and a field experiment involving health agencies with jurisdiction in 42 counties in the United States.

While public health officials periodically design relatively static health-promotion campaigns for both ongoing and new infections, social media platforms offer a living repository of messages that are generated by the community every day. This repository allowed us to produce the first automated, adaptive system to automatically gather and recommend HIV prevention and testing messages for counties in the United States, whereas previous approaches did not offer this level of automation and contextual specificity. Our computational processes were designed to continuously select HIV prevention and testing messages from US social media posts, thus using an inexpensive and creative way of finding the messages available for dissemination. We adopted a well-established concept in persuasive communications and behavioral interventions to guide the development of the computational processes. Specifically, *actionability* is defined as the probability that a message or intervention communicates what recipients might do, makes recommendations that are feasible for recipients, inspires recipients by describing what others do, or increases the sense that a behavior can be executed (3–5). The posts are then curated based on computational models of actionability, acceptability, and selected to also be appropriate for the priority HIV population of men who have sex with men (MSM). We chose the priority population of MSM first and foremost because they are central to the National End the HIV Epidemic Strategy (1) and thus of critical interest for county and state health departments tasked with HIV prevention and testing. In addition, MSM are difficult to reach with traditional health communication channels but receptive to new media technologies (6–10). Among MSM, those of color are also heavy users of social media (11), thus making these technologies particularly relevant to priority segments of MSM (12).

As government agencies and community-based organizations (CBOs) deploy HIV telehealth and home testing, enrolling individuals in remote HIV care and self-testing must necessarily rely on digital messages to recruit into telehealth platforms. Such methods are promising based on evidence about the behavioral impact of messages from past studies. First, communications containing behavioral-skills messages or arguments can increase condom use by 4% (OR = 1.82, reference: before the intervention containing behavioral-skills messages or arguments) compared with messages without that content (OR = 1.57, reference: before the intervention containing messages without that content) (13, 14). Second, moderate levels of campaign message exposure (i.e. seven times in the past year) have been shown to increase the likelihood of MSM testing for HIV by 96% and the likelihood of them using proper lubricant by 58% (15). Third, national media efforts such as the *Get Yourself Tested* campaign have shown to be effective at increasing HIV/STI prevention and testing behaviors among youth (16–21). Our method also tested the impact of selecting actionable messages, which is consistent with theories of behavioral change in which behaviors are predicted by clear intentions to execute a behavior (4, 22), recommendations that enable goal setting (23–27), and what other individuals do in a particular domain (28–30). Even though these messages should be highly frequent, as we show presently, they are only a minute fraction of the health-relevant messages that circulate on social media. Therefore, developing AI methods to identify them is potentially very important.

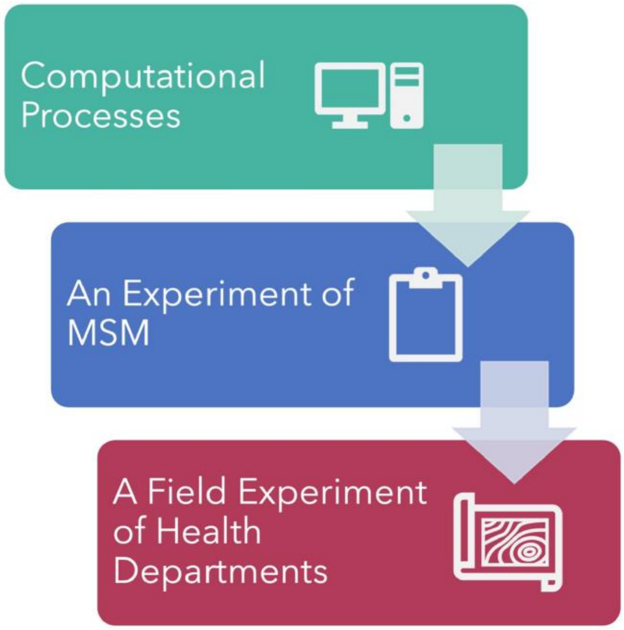

Fig. 1. The project's processes.

Our project had three phases (see Fig. 1). First, we developed the computational processes to select the most promising messages out of what communities posted on social media on any given day. Second, we conducted a preregistered experiment to determine if a sample of MSM perceived these posts as actionable, personally relevant, and effective, as hypothesized. This experiment included a manipulation of human vetting to determine how much this process contributed to selecting the most promising messages. Third, we conducted a preregistered field experiment in which, over a period of 17 months, 19 public health agencies and CBOs responsible for improving the health of 42 counties received daily message recommendations for a period of 8 weeks. We recommended 10,151 unique messages, including 5,154 experimental messages selected by our computational methods including human vetting and 4,997 control ones randomly picked from the remaining messages. We then analyzed which messages the agencies posted and the reception of these messages in their communities.

Results

In this project, we first developed a series of computational processes to automatically generate an HIV prevention and testing campaign for counties in the United States. The procedures involved an AI classifier and a set of keywords, hashtags, and engagement selections to make sure the messages were appropriate for MSM and for any county in the United States. We developed a training dataset with 13,314 HIV tweets obtained from keyword searches and expert accounts (e.g. federal agencies, nonprofit organizations, and HIV/STI researchers). Trained research assistants coded the training tweets based on three questions and achieved an inter-rater reliability of 0.78. The first question concerned the presence of a direct recommendation: *Does [the message] encourage either the audience or other people to act (or not act) in certain ways (does it tell people what they should/should not do)?* The second question concerned the presence of information that would allow the reader to enact the behavior: *Does [the message] provide information about an HIV-related behavior that is actionable to the audience?* The third question concerned the presence

Table 1. Performance of the actionability classification model using the validation dataset and the bi-LSTM model.

Metrics Validation dataset	Actionable ($n = 164$)	Nonactionable ($n = 1,161$)	Average
F1 score	0.78	0.96	0.87
Precision	0.69	0.98	0.84
Recall	0.90	0.94	0.92

	Top K					
	10	**30**	**50**	**90**	**110**	**1,000**
F1 score	1.00	0.97	0.96	0.90	0.79	0.76
Precision	1.00	0.97	0.96	0.90	0.89	0.66
Recall	1.00	0.97	0.96	0.90	0.89	0.88

The F1 measure is a single score that balances the concerns of both precision and recall in one number. Precision quantifies the number of positive predictions that belong to the positive cases, and recall quantifies the number of positive predictions made from all positive examples in the dataset.

of descriptions of the behavior of others: *Does [the message] directly describe other people's HIV-related behavior?* A message stating, "This is how you use a condom" and linking to a video showing the application of a condom was coded as actionable, as was a message stating, "50% of teenagers aren't using condoms. Condom works [*sic*]!." In contrast, messages such as, "Modeling shows that a well-implemented #AIDS vaccine could eliminate the majority of new #HIV infections" and "Condoms are safe" were coded as nonactionable. The initial training dataset involved 13,314 messages, out of which only 607 were actionable, supporting the need for methods to automate the selection.

This dataset was used to train a classifier using different machine learning techniques, including random forest, Bayes, logistic regression, support vector machines (SVM), long-short-term memory network (LSTM), and a bi-directional LSTM model (bi-LSTM). After a two-stage validation with the initial dataset of 13,314 messages and a second dataset containing 1,634 messages initially classified as actionable, the bi-LSTM model exhibited excellent performance. These analyses appear in Table 1, both for the classification of the 1,634 messages and for the top 1,000 messages.

We next conducted a preregistered online experiment using a within-subjects design to gauge the persuasiveness of the messages selected by the classifier. Two hundred and sixty MSM viewed 24 experimental messages selected to be top-ranked by the classifier and relevant based on keywords (see Materials and methods), as well as 12 control messages randomly selected from the HIV-related keyword search on X (formerly Twitter). Of the 24 messages selected by the classifier, 12 were vetted by a researcher, and 12 were not. After viewing each of the 36 messages, which were presented in random order, participants reported the degree to which they were actionable, appropriate for MSM, accurate, personally relevant, and effective, and the degree to which they were willing to share them online using a four-point scale, from 1 (*not at all*) to 4 (*definitely*), or a seven-point scale, from 1 (*strongly disagree*) to 7 (*strongly agree*). The study was approved by the University of Pennsylvania Institutional Review Board (IRB).

We performed mean comparisons to determine if the messages selected by the classifier were more actionable than those selected by the control method, in addition to gauging the impact of human vetting. Friedman rank-sum tests were conducted to estimate omnibus differences among three message conditions for each dependent measure (Table S2). These tests were supplemented with Wilcoxon tests for pairs of means, some of which appear in Fig. 2 (for the full set, see Supplementary material). As shown, relative to the control messages, the messages selected by the classifier were reported to be more actionable, appropriate for MSM, accurate, personally relevant, and effective, in addition to eliciting stronger sharing intentions. The advantage of the classifier was present irrespective of whether the messages were vetted by a human being, although vetted messages performed better than the unvetted ones. Compared with the unvetted messages, the vetted ones were deemed more actionable, appropriate for MSM, personally relevant, and effective, and elicited stronger sharing intentions as well.

Last, we launched a preregistered field experiment with public health agencies and CBOs in charge of HIV prevention and testing. We used a standardized protocol to recruit health departments and other agencies initially as participants and later as our collaborators. We approached about 100 agencies and preregistered a recruitment deadline of 2023 February 28. We recruited 19 institutions covering 42 counties, 2 of which dropped out from the study due to staffing issues. Health departments and CBOs completed a baseline questionnaire regarding their messaging about HIV prevention and testing on social media and were asked to participate for 8 weeks and to post between 5 and 10 messages from the total of ∼183 messages recommended each week. Agency posting is defined as sharing a recommended post with the addition of the agency's own commentary on social media platforms such as X (formerly Twitter) and Facebook. Two agencies did not post any messages on their social media accounts. At the end of the 8-week period, each participant completed the same questionnaire again. Health departments and CBOs were blind to our hypotheses and the process used to select messages and did not know what messages were experimental or control. The counties, states, and types of agencies we recruited appear in Table 2 and Fig. 3. Each agency designated a public health official or social media staff member to participate in the study, which lasted 8 weeks. The designated participant was also offered a social media assistant to help them with the selection and posting should they desire. Five participants accepted the social media assistant, who was blind to our hypotheses, did not know what messages were experimental or control, and followed the directions of the agency. The designated participant from each agency, who was also blind to the study hypotheses, signed an informed consent approved by the University of Pennsylvania IRB.

Agencies were offered an X app that automatically made recommendations (https://apps.apple.com/us/app/armt-hiv/id1532288199) as well as Buffer or Hootsuite, which are commercial social media management programs. Over a period of 8 weeks, each participating agency was recommended an average of 30 messages a day, 6 days a week, for a total of 1,461 recommendations (563 unique messages) per agency in the 8-week study period. Approximately half of these messages were experimental, and half were control messages. All agencies enrolled during the same period received the same set of recommended messages. As mentioned, the agency participant was asked to share between 5 and 10 unique recommended messages a week, using X, Facebook, or Instagram, depending on their preferences (see Fig. 4). Fifteen of the agencies used Twitter and four used Facebook. We recommended a total of 10,151 unique messages during the study period, which went from 2022 February 2 to 2023 June 24. We used the Twitter API v2 endpoint or Hootsuite or Buffer Analytics to extract the participating accounts' posts and their engagement metrics (i.e. retweet count, reply count, like count, and quote count) during the analysis period. We analyzed whether each experimental

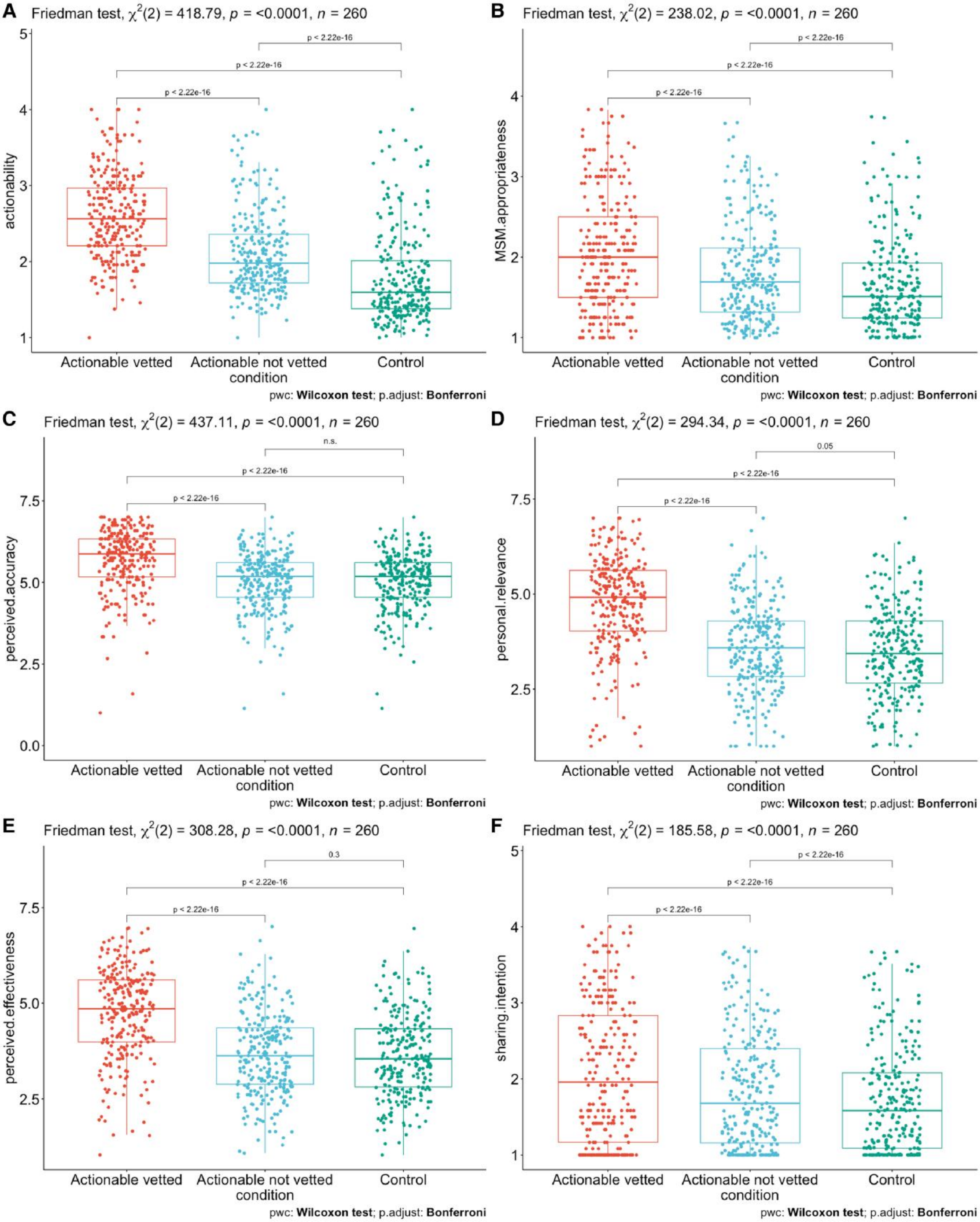

Fig. 2. Online experiment with MSM. Mean differences in perceived actionability (top left, A), appropriateness for MSM (top right, B), perceived accuracy (middle left, C), personal relevance (middle right, D), perceived effectiveness (bottom left, E), and sharing intentions (bottom right, F) across the three message conditions (actionable vetted, actionable not vetted, and HIV-related control).

and control message was posted by the agency as well as the actual engagements, including shares (e.g. retweets), quotes, replies, and likes of any posted messages. We repeated the engagement analyses using the average engagements as the outcome. The analyses followed an intent-to-treat model in which all agencies, including the four that dropped out or were deemed noncompliant, were included. Table S4 contains analyses without these accounts, which were very similar to the ones we report here.

The experimental messages were selected for agency posting 273 times relative to 39 times for the control ones. Collectively, they also received more replies than the control ones (replied to = 11 times, 4% of the agency posts, vs. 0 times, 0% of the agency posts, for experimental and control messages; reshared = 85 times, 31% of the agency posts, and 14 times, 36% of the control messages; liked = 143 times, 0.52 likes per posted message, vs. 37 times, 0.95 likes per posted message, for experimental and control messages; quotes = 2 times each for experimental and control messages; Table S3). The total number of engagements was 241 for experimental and 53 for control messages. When dividing each message's engagements by the total number of posts (i.e. 273 for experimental and 39 for control), the average number of engagements was $M = 6.47 \times 10^{-5}$ ($SD = 8.28 \times 10^{-4}$) for the experimental messages and $M = 1.03 \times 10^{-4}$ ($SD = 2.50 \times 10^{-3}$) for the control messages. These two numbers were very small because most recommended messages had 0 engagement but suggest that the experimental messages might have elicited less average engagement. Nonetheless, separate multilevel regression analyses clustered by agency showed no statistical differences in the average engagement between the experimental and control messages, $b = -0.00$, $P = 0.106$ and incidence-rate ratios = 0.62, $P = 0.733$.

As preregistered, analyses were conducted for agency posts and for the sum of reshares, replies, likes, and quotes for each agency. We performed logistic regressions for posting status (i.e. 1: posted and 0: not posted). We repeated linear regressions with the total engagements (i.e. the total number of replies, retweets, likes, and quotes) and, although not preregistered, for the average number of engagements (i.e. the total number of engagements divided by the total number of posts) as the outcomes. We also conducted zero-inflated negative binomial regression analyses that were not preregistered because of an excessive number of zeros in the actual and average engagement outcomes. We also assessed for evidence of clustering by user and date. The analyses of these data appear in Table 3. As shown, posting was 6.96 times more likely for the experimental than the control messages. In addition, the total number of engagements followed the greater posting of experimental than control messages, although the average number of engagements did not differ for experimental and control messages. The day when a message was recommended had a positive effect on the posting status but exerted no influence on the level of engagement. Also, the intraclass correlation coefficient of user handles and recommendation date revealed moderate to large clustering, but our experimental effects

Table 2. Counties, states, number, and type of agencies that participated in the field experiment.

Counties and states	Number of and types of agencies
Pueblo, CO	1 PHD
Wilmington, DE	1 CBO
Clayton, GA	1 PHD
Banks, Dawson, Forsyth, Franklin, Habersham, Hall, Hart, Lumpkin, Rabun, Stephens, Towns, Union, White, GA	1 PHD
Butts, Carroll, Coweta, Fayette, Heard, Lamar, Meriwether, Pike, Spalding, Troup, Upson, GA	1 PHD
Scott, IA	1 PHD
Champaign, IL	1 PHD
Hardin, LaRue, Marion, Meade, Nelson, Washington, KY	1 PHD
Madison, KY	1 PHD
Atlantic City, NJ	1 CBO
Hudson, NJ	1 CBO
Middlesex, NJ	1 CBO
Albany, NY	2 CBOs
Broome, NY	1 CBO
Cuyahoga, OH	1 PHD
Clatsop, OR	1 PHD
Philadelphia, PA	1 PHD, 1 CBO

PHD, public health department; CBO, community-based organization.

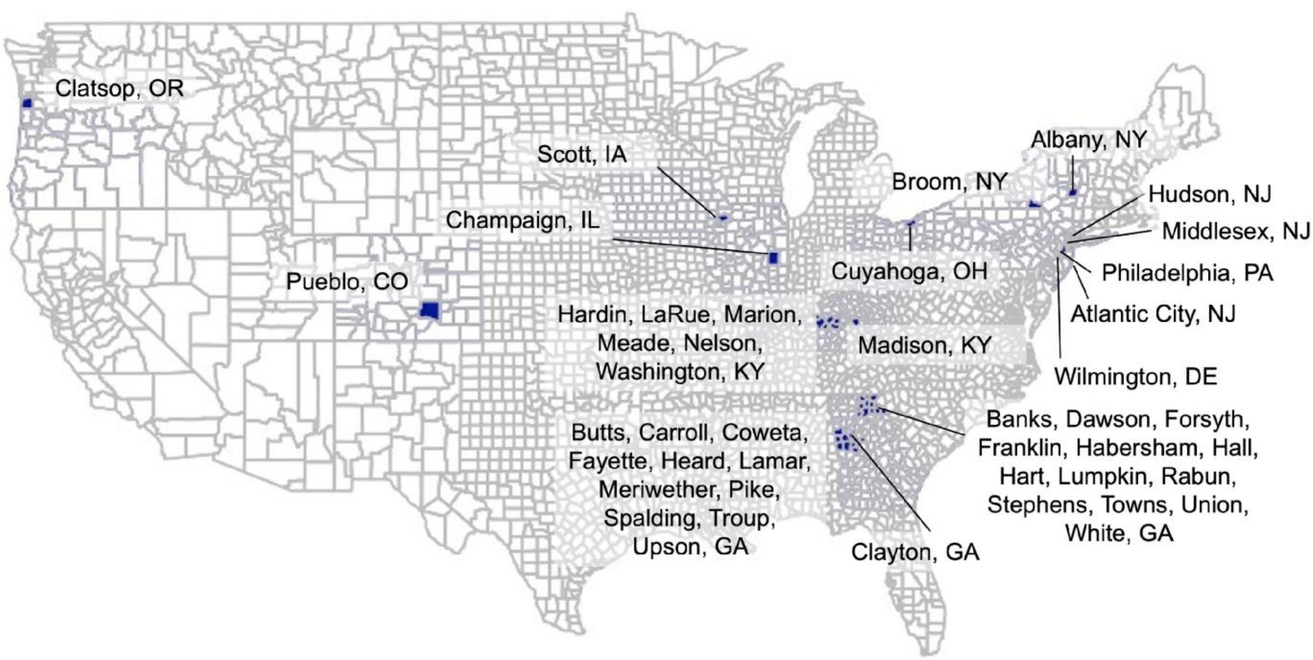

Fig. 3. Counties with participating agencies.

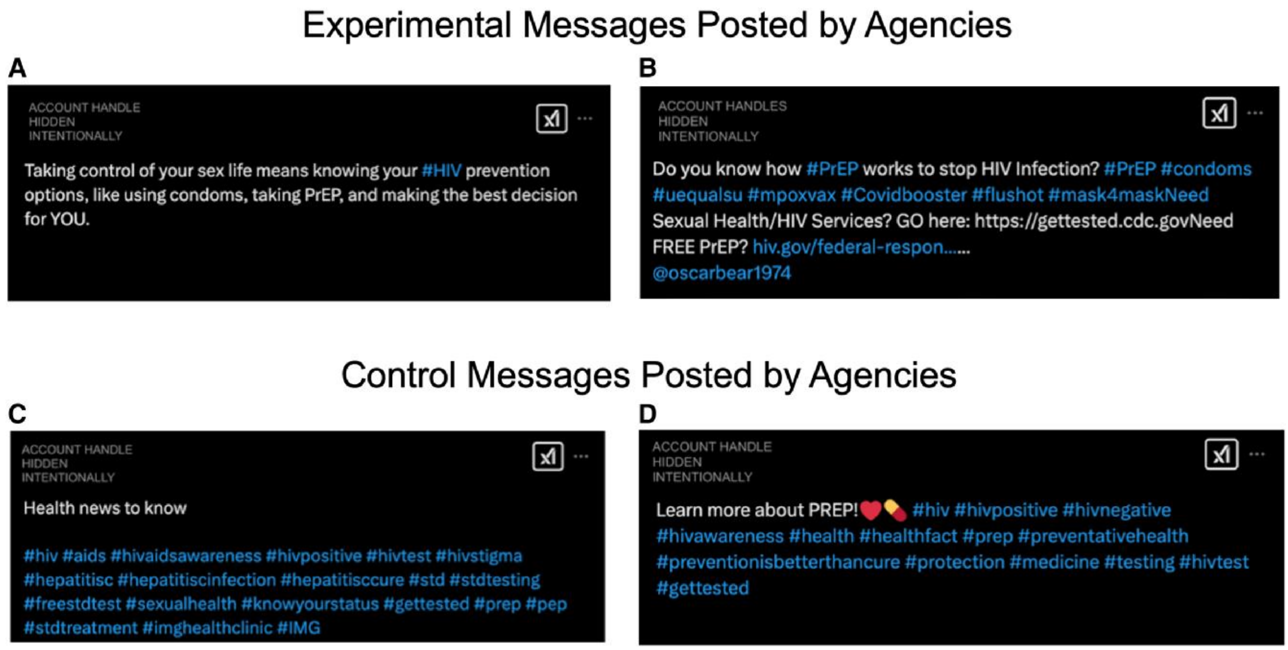

Fig. 4. Examples of experimental and control messages posted by agencies (top: experimental—A and B and bottom: control—C and D). Credit to the post creators: A) http://www.x.com/anyuser/status/1563904312614477825, B) http://www.x.com/anyuser/status/1650690310333235203, C) http://www.x.com/anyuser/status/1641060164521304064, and D) http://www.x.com/anyuser/status/1555276083989291008.

remained in analyses that clustered messages by handles and recommendation date.

Questionnaire responses collected at baseline and the end of the 8-week period (i.e. posttest) were used to gauge the reactions of the personnel who participated in the study. Participants reported posting more messages regarding HIV prevention and testing on social media at the posttest, $M = 5.00$, $SD = 0.58$, than at the baseline measurement, $M = 3.12$, $SD = 1.41$, $t(10) = 4.92$, $P = 0.001$. They also reported a reduction in the difficulties of identifying and/or creating relevant HIV messages to disseminate on social media [baseline: $M = 2.59$, $SD = 0.94$ and posttest: $M = 2.31$, $SD = 1.38$; $t(10) = 0.52$, $P = 0.614$], an increase in perceived effectiveness of the available HIV messages to be disseminated on social media [baseline: $M = 2.29$, $SD = 0.85$ and posttest: $M = 2.77$, $SD = 1.01$; $t(10) = 1.17$, $P = 0.267$], and an increase in satisfaction with the methods to disseminate relevant HIV messages on social media at the posttest than the baseline [baseline: $M = 2.41$, $SD = 0.71$ and posttest: $M = 2.85$, $SD = 1.28$; $t(10) = 0.67$, $P = 0.518$]. (Note that the comparisons over time were not preregistered but are provided to assist the reader.)

Discussion

Three studies showed the success of our computational method to automatically select decentralized, real-time messages that are perceived as actionable, personally relevant, and effective. The method had acceptable computational properties and led to selections that were both persuasive for a sample of MSM and conducive to posting by the public health agency and CBO staff of 42 US counties. This method also proved useful in a priority area for our nation and other countries with high social media penetration rates, generating solutions that will save time and resources in future efforts to improve health and potentially curb other urgent problems such as climate change.

As shown in the results of the online experiment, MSM rated actionable vetted messages as more persuasive and accurate than actionable not vetted ones. These findings suggest that a brief human vetting process, in addition to the classifier inference process, can select optimal messages without reducing the efficacy of our method. Algorithmic bias has become a growing concern in the use of machine learning in many areas (31–36), including health (37). Bias can be seeded in the process of developing machine learning models, particularly without a systematic approach to collecting human annotations. Despite the potential to ameliorate this bias through human annotation and expert opinions (e.g. adoption of a community advisory approach), we urge researchers to include an additional process to vet the processes and improve algorithms over time.

The present method can allow public health agencies and CBOs to reuse the low-hanging fruit of publicly available messages that are generated by a variety of community sources. Although an increasing number of agencies use social media for health communications, particularly following the COVID-19 pandemic, the numbers of HIV and other health messages disseminated at the county and state levels are still underwhelming (38). In our field experiment, only 42% of the public health agencies and CBOs reported posting messages on social media more than once a week at baseline, and the percentage jumped to 68% after the agencies used our methods. This substantial increase in posting demonstrates that these agencies and CBOs valued (see Appendix C for evaluations provided by the agencies) and benefited from the methods we examined. The present study offers the first empirical evidence for the successful automatic selection of messages for community dissemination. The proposed digital selection method using decentralized, real-time content can also allow health officials to systematically record the message selection and compare selection strategies in terms of audience engagement.

Like any research, the present work has some noteworthy limitations. First, the proposed work used social media messages obtained from X's (formerly Twitter) official API. The change in the official API plan will require updates to the program's data sources. Yet, as our method can process different short text messages

Table 3. Results of regressions and multilevel models with posting status and engagement after posting as the outcomes in the field experiment.

	Regression			MLM clustered by agency handles			MLM clustered by recommendation date		
Predictors	**Estimates**	**CI**	**P**	**Estimates**	**CI**	**P**	**Estimates**	**CI**	**P**
Logistic regression (posting status)									
(Intercept)	0.00	0.00–0.00	<0.001	0.00	0.00–0.00	<0.001	0.00	0.00–0.00	<0.001
Experimental condition	6.96	5.04–9.89	<0.001	7.02	5.03–9.79	<0.001	7.08	5.05–9.93	<0.001
Difference in recommendation (day)	1.00	1.00–1.01	<0.001	1.01	1.01–1.02	<0.001	1.01	1.00–1.01	<0.001
Random effects									
σ^2				3.29			3.29		
τ_{00}				3.25_{agency}			0.73_{date}		
ICC				0.50			0.18		
n				19_{agency}			369_{date}		
Observations	26,763			26,763			26,763		
R^2 Tjur	0.009			0.322/0.659			0.225/0.367		
AIC	3,158.9			3,040.3			3,115.2		
Log-likelihood	−1,576.5			−1,516.1			−1,553.6		
Linear regression (actual engagements)									
(Intercept)	0.00	−0.00, 0.01	0.376	−0.00	−0.01, 0.01	0.837	0.01	−0.00, −0.01	0.634
Experimental condition	0.01	0.01–0.02	<0.001	0.01	0.01–0.02	<0.001	0.01	0.01–0.02	<0.001
Difference in recommendation (day)	0.00	−0.00, 0.00	0.220	0.00	−0.00, 0.00	0.318	0.00	−0.00, 0.00	0.209
Random effects									
σ^2				0.03			0.03		
τ_{00}				0.00_{agency}			0.00_{date}		
ICC				0.01			0.00		
n				19_{agency}			369_{date}		
Observations	26,763			26,763			26,763		
R^2/R^2 adjusted	0.002/0.001			0.002/0.009			0.002/0.005		
AIC	−17,275.14			−17,359.50			−17,243.14		
Log-likelihood	8,641.568			8,684.749			8,626.569		
Zero-inflated negative binomial regression (actual engagements)									
Count model									
(Intercept)	0.09	0.01–0.52	0.007	0	0.00–0.04	<0.001	0	0.00–0.00	<0.001
Experimental condition	4.27	2.75–6.61	<0.001	6.05	3.84–9.53	<0.001	5.49	3.45–8.72	<0.001
Difference in recommendation (day)	1	1.00–1.00	0.804	1.01	1.00–1.01	0.041	1	1.00–1.00	0.237
(Intercept)	1.52			0.26			0.01		
Zero-inflated model									
(Intercept)	21.13	2.69–166.22	0.004	11.16	1.64–76.13	0.014	0	0.00–Inf	0.997
Random effects									
σ^2				5.11			6.01		
τ_{00}				3.34_{agency}			1.09_{date}		
ICC				0.39			0.15		
n				19_{agency}			369_{date}		
Observations	26,763			26,763			26,763		
Marginal R^2/conditional R^2	—			0.154/0.488			0.094/0.233		
AIC	2,328.6			2,183.619			2,326.891		
Log-likelihood	−1,159.3			−1,085.81			−1,157.445		
Linear regression (average engagements)									
(Intercept)	0.00	0.00–0.00	0.010	0.00	−0.00, 0.00	0.455	0.00	−0.00, 0.00	0.052
Experimental condition	−0.00	−0.00, 0.00	0.091	−0.00	−0.00, 0.00	0.107	−0.00	−0.00, 0.00	0.100
Difference in recommendation (day)	0.00	0.00–0.00	0.016	0.00	−0.00, 0.00	0.303	0.00	0.00–0.00	0.018
Random effects									
σ^2				0.00			0.03		
τ_{00}				0.00_{agency}			0.00_{date}		
ICC				0.01			0.00		
n				19_{agency}			369_{date}		
Observations	26,763			26,763			26,763		
R^2/R^2 adjusted	0.000/0.000			0.000/0.007			0.000/0.004		
AIC	−260,924.0			−260,983.1			−260,863.9		
Log-likelihood	130,466			130,496.5			130,436.9		
Zero-inflated negative binomial regression (average engagements)									
Count model									
(Intercept)	0	NA–NA	NA	0	0.00–0.00	<0.001	0	0.00–0.00	<0.001
Experimental condition	0.61	0.06–6.55	0.686	0.63	0.04–9.15	0.735	3.5	0.13–96.68	0.459
Difference in recommendation (day)	1	0.99–1.01	0.921	1	0.99–1.01	0.625	0.87	0.87–0.88	<0.001
(Intercept)	1.85E + 13			54,271,021.8			7.29E13		
Zero-inflated model									
(Intercept)	0	0.00–Inf	0.966	0	0.00–Inf	0.982	0	0.00–Inf	1
Random effects									
σ^2				10.13			8.33		
τ_{00}				0_{agency}			$11{,}205.77_{date}$		

(continued)

Table 3. Continued

Predictors	Regression			MLM clustered by agency handles			MLM clustered by recommendation date		
	Estimates	CI	P	Estimates	CI	P	Estimates	CI	P
ICC				—			1		
n				19_{agency}			369_{date}		
Observations	26,763			26,763			26,763		
Marginal R^2/Conditional R^2	—			0.016/NA			0.032/0.999		
AIC	NA			NA			NA		
Log-likelihood	NA			NA			NA		

Estimates, odds ratios for logistic regression analyses, unstandardized coefficients for linear regression analyses, and incidence-rate ratios for zero-inflated negative binomial regression analyses; CI, confidence interval; *P*, *P*-value; σ^2, residual variance; τ_{00}, random intercept variance; ICC, intraclass correlation coefficient; *N*, number of clusters; R^2 Tjur, pseudo-*R* squared values for multilevel models for binary outcomes; R^2, *R* squared; R^2 adjusted, *R* squared that has been adjusted for the number of predictors in the model; AIC, Akaike information criterion; NA, not available; Inf, Infinity. Agency posting status is a binary outcome, where 1 indicates that the agency has posted the recommended message with their commentary to their official account and 0 indicates they have not. Actual engagement is a continuous variable that sums all engagement metrics of the post (i.e. the total number of replies, retweets, likes, and quotes). Average engagement is a continuous outcome calculated by dividing total engagements by the number of messages the agency quoted.

without any other metadata, it is not limited to X or any particular social media platform. For instance, interested readers may access data through the Reddit API, the Meta Content Library, or the Content Library API. Additionally, they may subscribe to online social media data library services such as *Brandwatch*. Second, in the online experiment, we measured MSM's evaluations of the messages and their sharing intentions, but future studies could examine the effects on behavioral change among MSM. Third, we did not consider the interactive or cumulative effects of multiple messages, but future studies could investigate this issue and the impact of different properties of the messages. Finally, we developed the model using 2010–2018 data without incorporating a step/procedure to adopt new information. Therefore, the performance and applicability of the identified model could be susceptible to social and cultural changes, miss new features, and exhibit concept drift. We thus recommend reviewing the model identification process when new HIV prevention methods emerge and when epidemiological changes occur. The present study fills a critical gap for timely and efficient public communication in the areas of HIV, health, and potentially other domains. It showcases the feasibility of incorporating AI to generate health-promotion campaigns to address specific health domains (i.e. HIV testing and prevention in this study). The proposed system can inform policymakers and government and nongovernment agency officials who seek to promote health on social media, a space that has grown to reach large segments of the US public. Living health-promotion campaigns on social media are indispensable to achieving the ambitious and unprecedented goal of ending the HIV epidemic by 2030 in the United States and may also inspire future campaigns for the benefit of society and the individuals who live in it. Similar living and adaptive methods using decentralized content will also be essential when humanity faces the next, inevitable pandemic.

Materials and methods

Computational studies

Social media datasets

We developed a training dataset with HIV tweets obtained from keyword searches and expert accounts (e.g. federal agencies, nonprofit organizations, HIV/STI researchers). Trained research assistants coded the training tweets based on three questions and achieved an inter-rater reliability of 0.78. The first question concerned the presence of a direct recommendation: *Does [the message] encourage either the audience or other people to act (or not act) in certain ways (does it tell people what they should/should not do)?* The second question concerned the presence of information that would allow the reader to enact the behavior: *Does [the message] provide information about an HIV-related behavior that is actionable to the audience?* The third question concerned the presence of descriptions of the behavior of others: *Does [the message] directly describe other people's HIV-related behavior?* As an example, a message stating, "This is how you use a condom" and linking to a video showing the application of a condom was coded as actionable, as was a message stating, "50% of teenagers aren't using condoms. Condom works [*sic*]!." In contrast, messages such as, "Modeling shows that a well-implemented #AIDS vaccine could eliminate the majority of new #HIV infections" and "Condoms are safe" were coded as nonactionable.

We used two X (formerly Twitter) datasets (i.e. training data collected in 2010–2016, $n = 13{,}314$, and validation data collected in 2018, $n = 13{,}250$) to collect human annotations. A team of three research assistants and researchers from the research team coded each message after training to achieve high intercoder reliability. If a message received a "yes" in response to any of the questions, it was coded as actionable; otherwise, it was coded as nonactionable. The intercoder reliability coefficient Fleiss's kappa = 0.78, which was satisfactory.

Preprocessing for natural language analyses

The tweets were preprocessed by replacing the digits, URLs, and X handles with, and respectively; adding a prefix and suffix … surrounding each hashtag in the tweet; transforming all characters of the tweet content into lowercase except "I" in all capital letters (i.e. adding a prefix to the word with); and using a token for infrequent words based on the vocabulary matrix obtained from the training data.

Machine learning

We first used the training dataset to develop a classifier using different machine learning techniques, including Random Forest, Bayes, logistic regression, SVM, LSTM network, and a bi-LSTM model. Next, we used the validation dataset to refine the classification model.

Keyword and hashtag selection

Several keyword and hashtag selections were introduced to increase the relevance of the messages. First, to ensure that our target behaviors were captured, we applied specific keywords and hashtags about HIV testing, PrEP use, condom use, and safer sex talk (see Appendix A). Second, to increase the relevance of the messages to MSM, we filtered out messages about females, pregnancy, ladies, and sisters. Third, to increase the geographic relevance of the messages, we filtered out content about other countries and specific locations in the United States.

Ranking based on actionability and engagement

We then ranked the messages based on the actionability score and the engagement metric (i.e. the actual favorite counts on X) to select the top 200 for human vetting.

Human vetting process

An author then vetted this collection of messages based on their consistency with CDC recommendations. For example, "PrEP plus condom [*sic*] can give you the best protection. See where you can get free condom [*sic*]." would be selected, whereas "abstinence is the most effective for preventing HIV/STI, [*sic*] it's free" would not be selected.

Online experiment with MSM

Preregistration

The online experiment was preregistered on the Open Science Framework (https://osf.io/tc54g/). We detailed our hypotheses, study design, sampling plan, and analysis plan in advance. There was no deviation from our analysis plan.

Selection of messages

We collected three sets of tweets, including actionable vetted, actionable not vetted, and control tweets, for the online experiment. The first two sets of tweets were collected based on keywords using the Twitter Streaming API and from HIV expert handles, including individual researchers, physicians, and governmental and nongovernmental health organizations ($n = 363$; see Appendix B) using the Twitter REST API v1.1. Figure S1 presents in detail the steps used to collect and select messages for the study. We applied the bi-LSTM classifier to infer the level of actionability (as a continuous score). We next ranked the tweets by the actionability scores, followed by ranking them by their engagement metric (i.e. the actual favorite counts on X to gauge acceptability) to select the top 200 messages. This procedure allowed us to select primarily based on actionability but to choose the more popular message when two messages had the same actionability score. We vetted messages out when they contained explicit content, profane language, or specific event details, and chose the first 24 messages to be the actionable vetted ones. Twenty-four messages were randomly selected from the pool of unvetted selected messages to be the actionable not vetted ones. Lastly, 24 messages were randomly selected from the HIV-related keyword search results to be included in the control message condition.

Participants

We recruited respondents from the survey participant platform Prolific in September and October 2023. The study advertisements indicated that the study would ask questions about men's sexual health and wellness and take ∼30 min to complete. Each respondent received $10 for their time upon completing the study. Three hundred fifty-four participants started the study, and 340 participants completed it. Per our preregistered exclusions, we excluded participants who failed the attention check items (i.e. $n = 80$), although including them did not alter the results. Our final dataset included 260 individuals: mean age = 36.73 years (SD = 12.15), 96% male, 4% male and transgender man, 18% Black, and 69% White (see Table S1).

Experimental procedures

The study used a within-subjects design. Participants gave their informed consent first and then answered questions to determine their eligibility to participate in the study. The screening items included questions about their age, country of residence, self-identified gender, sex at birth, history of having oral/anal sex with a man, and sexual orientation. Participants were eligible to continue if they were male, at least 18 years of age, indicated having oral or anal sex with another man in their lifetime, self-identified as gay or bisexual, and currently residing in the United States. After responding to a short eligibility questionnaire, participants read 36 randomly selected messages (12 actionable vetted messages, 12 actionable not vetted messages, and 12 control messages) in a random order to minimize the order effect. The message display resembles the actual viewing experience on X. Participants were asked to take their time reading through the posts, and they rated statements for six variables after viewing each post (i.e. message actionability, message appropriateness for MSM, perceived accuracy, personal relevance, perceived effectiveness, and sharing intention). After viewing and rating all 36 posts, participants were then thanked for their participation and debriefed.

Measures

Message actionability

Participants rated two items on a four-point scale (i.e. definitely [4], quite a bit [3], somewhat [2], not at all [1]) designed to measure how actionable the message was: "This message gives the audience a clear indication of what to do about HIV prevention or testing," and "The message points to concrete resources that may help to implement HIV prevention or testing." These two items were combined into an index, with a higher score indicating greater message actionability ($r = 0.93$, $P < 0.001$, $M = 2.16$, SD = 0.54, see Table S1).

Message appropriateness for MSM

Participants rated one item on a four-point scale (i.e. definitely [4], quite a bit [3], somewhat [2], not at all [1]) designed to measure how appropriate the message was for MSM: "This message is designed specifically for gay or bisexual men." A higher score indicates greater message appropriateness for MSM ($M = 1.84$, SD = 0.61).

Perceived accuracy was based on one item: "I believe this message is truthful." This item was answered on a seven-point scale (i.e. strongly agree [7], agree [6], somewhat agree [5], neither agree nor disagree [4], somewhat disagree [3], disagree [2], strongly disagree [1]). The mean of this item indicates the perceived message accuracy ($M = 5.25$, SD = 0.87).

Personal relevance was based on two items: "This message seems to have been designed for the benefit of people like me," and "This message has the potential to change the behavior of people like me." Each item was answered on a seven-point scale (i.e. strongly agree [7], agree [6], somewhat agree [5], neither agree nor disagree [4], somewhat disagree [3], disagree [2], strongly disagree [1]). The mean of these two items was used to indicate the perceived message relevance to participants ($r = 0.92$, $P < 0.001$, $M = 3.95$, SD = 1.11).

Perceived effectiveness was based on one item: "This message has the potential to change my behavior." This item was answered on a seven-point scale (i.e. strongly agree [7], agree [6], somewhat agree [5], neither agree nor disagree [4], somewhat disagree [3], disagree [2], strongly disagree [1], $M = 4.02$, $SD = 1.04$).

Sharing intention

We also included one item on a four-point scale (i.e. definitely [4], quite a bit [3], somewhat [2], not at all [1]) to measure the extent to which participants would share the message with their social network: "I would share this message with my network if I encountered it on social media." A higher score indicates a stronger intention to share the message ($M = 1.88$, $SD = 0.77$).

Field experiment

Preregistration

We preregistered on the Open Science Framework (https://osf.io/mbjxs). We detailed our hypotheses, study design, sampling plan, and analysis plan in advance. We updated three details on the preregistration plan on 2022 November 28: (i) Changing the target sample from local health departments to any local health organizations to facilitate our recruitment goals, (ii) updating the estimated recruitment end date from 2022 August 31 to 2023 February 28, and (iii) adding other social media management tools such as Buffer in addition to Hootsuite. We expected no effects of these changes on the study results.

Continuous selection of messages

As in the previous experiment, the tweets were collected via HIV keyword searches using the Twitter Streaming API and by crawling the posts from 363 HIV expert handles, including individual researchers, physicians, and governmental and nongovernmental health organizations (see Appendix B) using the Twitter REST API v1.1. This procedure was applied daily and yielded an average of over 180,000 tweets a day. Following the same selection procedure described in the online experiment with MSM (Fig. S1), we first applied the bi-LSTM classifier to rate each tweet's actionability as a continuous score. Next, we ranked the tweets by the actionability scores, followed by ranking them by their engagement metric (i.e. the actual favorite counts on X to gauge acceptability) to select the top 200 messages. This procedure allowed us to select primarily based on actionability but to choose the more popular tweet when two tweets had the same actionability score.

As in the previous experiment, the experimental tweets were selected from the top 200 messages through a human vetting process that involved a researcher going down the ranked list after some tweets were vetted out due to containing explicit content, profane language, or specific event details (e.g. a condom distribution event relevant to a particular location). The control tweets were selected through a random selection from the remaining pool of the top 200 messages. We then recommended 30–40 messages a day 6 days a week for 2 months to each agency. Half of the recommended messages were experimental, and half were control. Overall, the experimental messages were more actionable than the control ones (experimental: $M = 2.78$, $SD = 0.51$; control: $M = 1.71$, $SD = 0.70$; $t(9{,}764) = 91.47$, $P < 0.001$), but the two groups of messages did not differ in actual favorite counts (experimental: $M = 16.50$, $SD = 862.23$; control: $M = 0.76$, $SD = 19.42$; $t(4{,}890) = 1.27$, $P = 0.202$). Each agency then curated and posted some of the recommended messages to their social media accounts weekly, based on their social media management schedules. All these messages were modified by the agencies, which added their own commentary, thoughts, or information through text, photos, videos, or URLs (i.e. "quote" messages that clearly indicate the original author).

Design and procedures

All agencies gave informed consent and completed a set of questionnaires at baseline. The questionnaire included questions asking about their current social media activities with respect to posting about HIV prevention and testing messages. Respondents answered the question, "In a typical month, how frequently do you post and/or repost health-promotion messages on social media surrounding HIV prevention/testing, condom use, and PrEP?" using a six-point scale (i.e. more than once a day [6], once a day [5], 1–2 times a week [4], once every week [3], once every 2 weeks [2], less than once a month [1]) and "In your work, how difficult is it to identify and/or create health-promotion messages to disseminate on social media surrounding HIV prevention/testing, condom use, and PrEP?," "To what extent do you feel that the messages you disseminated on social media have been effective at promoting HIV prevention/testing, condom use, and PrEP use in your community?," and "To what extent are you satisfied with the current methods you use to disseminate health-promotion messages on social media surrounding HIV prevention/testing, condom use, and PrEP?" using a five-point scale (i.e. a great deal [5], a lot [4], a moderate amount [3], a little [2], not at all [1]). At the end of the baseline questionnaire, they indicated whether they wanted to participate in the 8-week study through (i) using a mobile application to connect their X, formerly Twitter, accounts to the application that disseminates recommended messages, (ii) using a social media management platform (e.g. Hootsuite or Buffer), (iii) using a social media management platform plus a social media coordinator program, or (iv) joining a social media coordinator program. The research team then set up respective participation options with the agencies. At the end of the study period, all public health agencies and CBOs completed the same questionnaire again and were compensated with US $200 and thanked for their participation.

Recording of dependent measures

We used the Twitter API v2 endpoint and Hootsuite Analytics (or other social media management platforms) to extract the agency posts and the engagement metrics for each post (i.e. retweet count, reply count, like count, and quote count) during the analysis period (i.e. 2022 February 2–2023 June 24). Both X and Facebook/Instagram accounts were tracked.

We compared each message posted on a public health agency or CBO account with our message database to determine if a particular message was recommended by our team via the ARMT-HIV App or the social media management program. Each user account (participant) contributed to, on average, 1,338 records ($SD = 220$) of posting status and engagement metrics in the final dataset used in analysis.

Full List of the Authors and Affiliations

Sarah Joseph, Pueblo Department of Public Health and Environment in Colorado, Colorado, USA.

Anthony Hron, Clayton County Health District, Georgia, USA.

Janet Davis, Georgia District 2 Public Health Departments, Georgia, USA.

Tito Terry, Georgia District 4 Public Health Departments, Georgia, USA.

Tiffany Peterson, Scott County Health Department in Iowa, Iowa, USA.

Corey Herrman, Champaign-Urbana Public Health District in Illinois, Illinois, USA.

Melissa Phillips, Lincoln Trail District Health Department in Kentucky, Kentucky, USA.

Jennifer Osborne, Lincoln Trail District Health Department in Kentucky, Kentucky, USA.

Kelley G. McBride, Madison County Health Department in Kentucky, Kentucky, USA.

Martin Hensley, Madison County Health Department in Kentucky, Kentucky, USA.

Adriana Todorov, New Jersey Public Health Association, New Jersey, USA.

Alain Morrissette, Albany Damien Center, New York, USA.

Georgett Watson, South Jersey AIDS Alliance, New Jersey, USA.

Ethan Knox, STAP Inc (Southern Tier AIDS Program), New York, USA.

Erin Lark, Cuyahoga County Board of Health in Ohio, Ohio, USA.

Elisa Long, Clatsop County Public Health Department in Oregon, Oregon, USA.

Carolina Guerrero-Lara, Clatsop County Public Health Department in Oregon, Oregon, USA.

Timothy Rissel, Action Wellness, Pennsylvania, USA.

Michele Raymond, Action Wellness, Pennsylvania, USA.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

National Science Foundation grant 2031972 (D.A.), National Institutes of Health grant R01MH094241 (D.A.), National Institutes of Health grant R01MH114847 (D.A.), National Institutes of Health grant DP1DA048570 (D.A.), National Institutes of Health grant R01AI147487 (D.A. and M.S.C.), National Institutes of Health grant P30AI045008 (M.S.C.).

Author Contributions

Conceptualization: D.A. and T.S.; Data curation: D.A. and M.S.C.; Formal analysis: M.S.C.; Methodology—online experiment: D.A., T.S., and M.S.C.; Methodology—field experiment: D.A., M.S.C., and H.J.; Methodology—analyses and computational infrastructure: M.S.C.; Methodology—coding and training dataset in the computational work: S.L.; Methodology—classification: C.Z., A.M., and M.S.C.; Investigation—agency recruitment: D.A., A.Z., and D.O.K.; Investigation—human vetting: M.S.C., A.S., and M.R.D.; Investigation—critical comment on the content generated by the classifier, leading to the addition of keyword searches: M.R.D. and D.A.; Project administration: M.S.C.; Software: A.M. and M.S.C.; Supervision: D.A. and T.S.; Validation: M.S.C.; Visualization: M.S.C.; Funding acquisition: D.A., T.S., and M.S.C.; Writing—original draft: D.A.; Writing—methods and results of the studies: M.S.C.; Writing—review and editing: All authors.

Data Availability

The data and R codes that support the findings of this study are available in OSF.io at https://osf.io/tc54g/?view_only=3a1e5d7703604359bc6949e2f99c5adf and https://osf.io/e5kdt/?view_only=e74e674bfae542229187e02b2be07ede. These links will be public after the manuscript is accepted for publication.

References

1 Fauci AS, Redfield RR, Sigounas G, Weahkee MD, Giroir BP. 2019. Ending the HIV epidemic: a plan for the United States. *JAMA*. 321(9):844–845.

2 Athey S, Grabarz K, Luca M, Wernerfelt N. 2023. Digital public health interventions at scale: the impact of social media advertising on beliefs and outcomes related to COVID vaccines. *Proc Natl Acad Sci U S A*. 120:e2208110120.

3 Cohen JB, Eduardo BA. 2018. The ADF framework: a parsimonious model for developing successful behavior change interventions. *J Mark Behav*. 3:81–119.

4 Albarracín D. *Action and inaction in a social world: predicting and changing attitudes and behavior*. Cambridge University Press, Cambridge, 2021.

5 Albarracín D, Sunderrajan A, Dai W. Action, inaction, and actionability: definitions and implications for communications and interventions to change behaviors. In: Hope D, Bevins R, editors. *Change and maintaining change. Nebraska Symposium on Motivation*. Vol. 65, Springer, Cham, 2018. p. 75–99.

6 Bauermeister JA, *et al*. 2015. Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the get connected! Program. *AIDS Behav*. 19:1860.

7 Allison S, *et al*. 2012. The intersection of youth, technology, and new media with sexual health: moving the research agenda forward. *J Adolesc Health*. 51:207–212.

8 Bauermeister JA, *et al*. 2015. Where you live matters: structural correlates of HIV risk behavior among young men who have sex with men in Metro Detroit. *AIDS Behav*. 19: 2358–2369.

9 Bauermeister JA, *et al*. 2015. The use of mystery shopping for quality assurance evaluations of HIV/STI testing sites offering services to young gay and bisexual men. *AIDS Behav*. 19: 1919–1927.

10 Pew Research Center. *Social media fact sheet*. Pew Research Center, 2021 [accessed 2023 Dec 19]. https://www.pewresearch.org/internet/fact-sheet/social-media/.

11 Smith A, Brenner J. *Twitter use 2012*. Pew Research Center, 2012 [accessed 2023 Dec 19]. https://www.pewresearch.org/internet/2012/05/31/twitter-use-2012/.

12 Chan MS, *et al*. 2021. Estimating the influence of Twitter on pre-exposure prophylaxis use and HIV testing as a function of rates of men who have sex with men in the United States. *AIDS*. 35: S101–S109.

13 Albarracin D, *et al*. 2003. Persuasive communications to change actions: an analysis of behavioral and cognitive impact in HIV prevention. *Health Psychol*. 22:166–177.

14 Albarracín D, *et al*. 2005. A test of major assumptions about behavior change: a comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. *Psychol Bull*. 131(6):856–897.

15 Flowers P, McDaid LM, Knussen C. 2013. Exposure and impact of a mass media campaign targeting sexual health amongst Scottish men who have sex with men: an outcome evaluation. *BMC Public Health*. 13:737.

16 Eastman-Mueller HP, Habel MA, Oswalt SB, Liddon N. 2019. Get Yourself Tested (GYT) campaign: investigating campaign awareness and behaviors among high school and college students. *Health Educ Behav*. 46:63–71.

17 Roston A, Suleta K, Stempinski K, Keith L, Patel A. 2015. Get Yourself Tested 2011–2012: findings and prevalence of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* at an urban public health system. *Int J STD AIDS*. 26:322–328.
18 McFarlane M, *et al.* 2015. GYT: get yourself tested campaign awareness: associations with sexually transmitted disease/HIV testing and communication behaviors among youth. *Sex Transm Dis*. 42:619–624.
19 Habel MA, *et al.* 2015. Does your college campus GYT? Evaluating the effect of a social marketing campaign designed to raise STI awareness and encourage testing. *Cases Public Health Commun Mark*. 8:51–70.
20 Garbers S, *et al.* 2016. Adapting the get yourself tested campaign to reach Black and Latino sexual-minority youth. *Health Promot Pract*. 17:739–750.
21 Friedman AL, *et al.* 2014. An assessment of the GYT: Get Yourself Tested campaign: an integrated approach to sexually transmitted disease prevention communication. *Sex Transm Dis*. 41: 151–157.
22 Fishbein M, Ajzen I. *Predicting and changing behavior: the reasoned action approach*. Psychology Press, London, 2011.
23 Avishai A, Conner M, Sheeran P. 2019. Setting realistic health goals: antecedents and consequences. *Ann Behav Med*. 53(12): 1020–1031.
24 Oettingen G, Gollwitzer PM. 2000. Goal setting and goal implementation. *Z Psychol*. 208:406–430.
25 Locke EA, Latham GP. *A theory of goal setting & task performance*. Prentice-Hall, Inc., 1990.
26 Doran GT. 1981. There's a S.M.A.R.T. Way to write management's goals and objectives. *Manage Rev*. 70:35–36.
27 SAMHSA Native Connections. *Setting goals and developing specific, measurable, achievable, relevant, and time-bound objectives*. n.d. https://www.samhsa.gov/sites/default/files/nc-smart-goals-fact-sheet.pdf.
28 Cialdini RB, Goldstein NJ. 2004. Social influence: compliance and conformity. *Annu Rev Psychol*. 55:591–621.
29 Cialdini RB, Trost M. Social influence: social norms, conformity, and compliance. In: Gilbert DT, Fiske ST, Lindzey G, editors. *The handbook of social psychology*. 4th ed. McGraw-Hill, New York, 1998. p. 151–192.
30 Bicchieri C. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press, Cambridge, 2006.
31 Robertson KK, Khoo C, Song Y. *To surveil and predict: a human rights analysis of algorithmic policing in Canada*. University of Toronto, Toronto, 2020.
32 Kordzadeh N, Ghasemaghaei M. 2022. Algorithmic bias: review, synthesis, and future research directions. *Eur J Inf Syst*. 31: 388–409.
33 Igoe KJ. *Algorithmic bias in health care exacerbates social inequities—how to prevent it*. Health Information Technology, 2021.
34 Mittermaier M, Raza MM, Kvedar JC. 2023. Bias in AI-based models for medical applications: challenges and mitigation strategies. *NPJ Digit Med*. 6:113.
35 Obermeyer Z, Powers B, Vogeli C, Mullainathan S. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science (1979)*. 366:447–453.
36 Nazer LH, *et al.* 2023. Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digit Health*. 2:e0000278.
37 Igoe KJ. 2021. Algorithmic bias in health care exacerbates social inequities—how to prevent it. Executive and Continuing Professional Education | Harvard T.H. Chan School of Public Health [accessed 2023 Mar 28]. https://www.hsph.harvard.edu/ecpe/how-to-prevent-algorithmic-bias-in-health-care/.
38 Balogun BA, Hogden A, Kemp N, Yang L, Agaliotis M. 2023. Public health agencies' use of social media for communication during pandemics: a scoping review of the literature. *Osong Public Health Res Perspect*. 14:235–251.