

Original Paper

Predicting Participant Engagement in a Social Media–Delivered Lifestyle Intervention Using Microlevel Conversational Data: Secondary Analysis of Data From a Pilot Randomized Controlled Trial

Ran Xu¹, PhD; Joseph Divito¹, BS; Richard Bannor¹, MPH; Matthew Schroeder², MS; Sherry Pagoto¹, PhD

¹Department of Allied Health Sciences, Institute for Collaboration in Health, Interventions, and Policy, University of Connecticut, Storrs, CT, United States

²Center for Aging Research, Indiana University, Indianapolis, IN, United States

Corresponding Author:

Ran Xu, PhD

Department of Allied Health Sciences

Institute for Collaboration in Health, Interventions, and Policy

University of Connecticut

Koons Hall, Room 326

Storrs, CT, 06269

United States

Phone: 1 860 486 2945

Email: ran.2.xu@uconn.edu

Abstract

Background: Social media–delivered lifestyle interventions have shown promising outcomes, often generating modest but significant weight loss. Participant engagement appears to be an important predictor of weight loss outcomes; however, engagement generally declines over time and is highly variable both within and across studies. Research on factors that influence participant engagement remains scant in the context of social media–delivered lifestyle interventions.

Objective: This study aimed to identify predictors of participant engagement from the content generated during a social media–delivered lifestyle intervention, including characteristics of the posts, the conversation that followed the post, and participants' previous engagement patterns.

Methods: We performed secondary analyses using data from a pilot randomized trial that delivered 2 lifestyle interventions via Facebook. We analyzed 80 participants' engagement data over a 16-week intervention period and linked them to predictors, including characteristics of the posts, conversations that followed the post, and participants' previous engagement, using a mixed-effects model. We also performed machine learning–based classification to confirm the importance of the significant predictors previously identified and explore how well these measures can predict whether participants will engage with a specific post.

Results: The probability of participants' engagement with each post decreased by 0.28% each week ($P < .001$; 95% CI 0.16%–0.4%). The probability of participants engaging with posts generated by interventionists was 6.3% ($P < .001$; 95% CI 5.1%–7.5%) higher than posts generated by other participants. Participants also had a 6.5% ($P < .001$; 95% CI 4.9%–8.1%) and 6.1% ($P < .001$; 95% CI 4.1%–8.1%) higher probability of engaging with posts that directly mentioned weight and goals, respectively, than other types of posts. Participants were 44.8% ($P < .001$; 95% CI 42.8%–46.9%) and 46% ($P < .001$; 95% CI 44.1%–48.0%) more likely to engage with a post when they were replied to by other participants and by interventionists, respectively. A 1 SD decrease in the sentiment of the conversation on a specific post was associated with a 5.4% ($P < .001$; 95% CI 4.9%–5.9%) increase in the probability of participants' subsequent engagement with the post. Participants' engagement in previous posts was also a predictor of engagement in subsequent posts ($P < .001$; 95% CI 0.74%–0.79%). Moreover, using a machine learning approach, we confirmed the importance of the predictors previously identified and achieved an accuracy of 90.9% in terms of predicting participants' engagement using a balanced testing sample with 1600 observations.

Conclusions: Findings revealed several predictors of engagement derived from the content generated by interventionists and other participants. Results have implications for increasing engagement in asynchronous, remotely delivered lifestyle interventions,

which could improve outcomes. Our results also point to the potential of data science and natural language processing to analyze microlevel conversational data and identify factors influencing participant engagement. Future studies should validate these results in larger trials.

Trial Registration: ClinicalTrials.gov NCT02656680; <https://clinicaltrials.gov/ct2/show/NCT02656680>

(*JMIR Form Res* 2022;6(7):e38068) doi: [10.2196/38068](https://doi.org/10.2196/38068)

KEYWORDS

weight loss; social media intervention; engagement; data science; natural language processing; NLP; social media; lifestyle; machine learning; mobile phone

Introduction

Background

Obesity is prevalent in the United States and is a known risk factor for cardiovascular disease, type 2 diabetes [1,2], and cancer. Although lifestyle interventions are effective for weight loss and diabetes prevention [3], they require numerous clinic visits for up to a year, which is burdensome for many people. Technology-delivered lifestyle interventions, by not requiring visits, are less burdensome for participants and have shown promising weight loss outcomes [4]. Some technology-based interventions use popular commercial social media platforms such as Facebook in an effort to meet people where they are [5,6]. Many social media users already use these platforms to discuss their health experiences [7,8]. Community-building features on social media platforms, such as private groups [9,10], make them particularly amenable to delivering group-based lifestyle interventions.

Systematic reviews and meta-analyses show support for the efficacy of social media-delivered lifestyle interventions [4,11]; however, this area of research is still nascent. Participant engagement, defined as posts in the group, replies to posts, “likes,” and votes in polls, appears to be an important predictor of outcomes [12-15]. For example, a study found that every 10 posts by participants corresponded to -0.5% weight loss [16]. Another study found only certain types of engagement predicted weight loss [17]. Interestingly, the degree of participant engagement reported in studies of social media-delivered interventions is highly variable, ranging from an average of once during the entire intervention period to 11 times per week [17-19]. Engagement also generally declines over time in these programs [16]. Our understanding of the factors that influence participant engagement in these interventions is limited. Emerging evidence in the web-based communication literature shows the importance of multilevel factors influencing participant engagement, such as the characteristics of posts (eg, post length and topic and popularity of the poster), characteristics of the conversation thread in response to the post (eg, sentiment and reciprocity behavior), and participant characteristics (eg, motivation and habits) [14,20-27]. However, research is scant in the context of social media-delivered behavioral interventions [28]. Furthermore, characteristics of the conversation thread (ie, other peoples’ replies and comments) preceding each participant’s engagement is often ignored, which could be valuable in terms of providing microlevel contextual data that influences each participant’s decision to engage.

A promising approach to increase our understanding of the factors influencing participants’ engagement in social media-based behavioral interventions is to study the content and interactions generated by the interventionists and participants during the intervention using natural language processing (NLP). Data collected directly from web-based platforms (eg, Facebook) can provide detailed, real-time behavioral information over the course of intervention programs. NLP can handle a large quantity of text, generate reliable qualitative coding [29], be leveraged to derive various real-time microlevel insights concerning the characteristics of the posts (eg, topics) and conversations that followed the posts (eg, sentiment), and understand how they affect participants’ decisions of engagement independently and aggregately. This has potential implications for the design and implementation of future interventions to increase participant engagement, which could lead to more favorable weight loss outcomes.

Objective

Using data from a 16-week pilot feasibility randomized weight loss trial that delivered lifestyle interventions via Facebook, drawing on multilevel factors influencing participant engagement identified by previous web-based communication literature, we derived various factors from the content generated by participants and interventionists over the course of the intervention, including characteristics of the posts (eg, poster, time, and topic), conversations that followed the post (eg, sentiment and receiving replies), and participants’ previous engagement behaviors, and assessed how well these factors predict participant engagement individually and all together in the context of a social media-delivered lifestyle intervention.

Methods

Study Design, Settings, and Participants

In a pilot feasibility randomized trial, we randomized 80 participants who were overweight or obese into 1 of 2 remotely delivered lifestyle interventions. We recruited people interested in losing weight via web-based advertisements at the University of Connecticut on ResearchMatch and in yard sale or neighborhood Facebook groups in 37 states across the United States between June and October of 2019. Inclusion criteria included BMI between 27 and 45 kg/m², smartphone ownership, active Facebook user (ie, comments or posts more than once a week), aged 18 to 65 years, and having daily internet access. Exclusion criteria included pregnancy or planning to become pregnant during the study, bariatric surgery or plans for bariatric surgery during the study, ≥5% weight loss in the past 3 months,

pre-existing conditions that preclude physical activity or dietary changes, taking medications affecting weight, incapable of walking one-fourth of a mile unaided without stopping, type 1 or type 2 diabetes, and participation in prior weight loss studies under the principal investigator.

Participants completed an orientation webinar before randomization to learn more about the study, and those still interested in participating were mailed a Wi-Fi scale (FitBit Aria, FitBit Inc) and asked to provide the staff with their log-in information for the scale so that weights could be recorded for the assessments. We randomized 80 participants to the 2 conditions.

Intervention Conditions

Overview

Participants were randomized to either a Facebook group in which new participants were continually enrolled during weeks 1 to 8 (open enrollment) or a Facebook group that included only the original 40 randomized participants (closed enrollment). In the open enrollment condition, 54 additional participants were enrolled between weeks 1 to 8 for a final group size of 94. However, we only included the original 80 randomized participants in this study to ensure all participants had an equal amount of time to engage in all 16 weeks of the intervention.

Facebook-Delivered Lifestyle Intervention

Both conditions received the identical 16-week lifestyle intervention based on the Diabetes Prevention Program (DPP) but modified to be delivered in a private Facebook group where twice-daily posts guided participants through the program, which was led by a dietitian (counselor) who was assisted by a student counselor. We adapted the DPP content to be appropriate for a web-based setting, as described elsewhere [30]. We gave each participant an individualized calorie goal that would facilitate a 1 lbs to 2 lbs weight loss weekly and asked them to use MyFitnessPal to track their calories daily. In addition, we asked participants to have the counselor review at least 2 weeks of their MyFitnessPal records, although they could request more as needed. The Facebook group was private such that only those invited by the study team could join the group and view the intervention content. We gave participants diet and exercise goals for the week each Monday and asked them to report progress on their goals in a conversation thread on Sunday and report their weight change for the week in a conversation thread

each Friday. In between, intervention posts addressed the learning objectives of each module of the DPP. The dietitian leading the group was instructed to reply to all participant posts and comments that merited a reply and otherwise hit a reaction (eg, like and laugh) button to acknowledge the participants' comments. Participants were encouraged to post to the group and reply to each other. The details of the intervention, study procedures, and primary study results can be found elsewhere [31].

Ethics Approval

This pilot feasibility randomized trial was approved by the University of Connecticut Institutional Review Board (H17-215) in October 2017.

Measures

Overview

We included all posts and comments or replies within posts from interventionists and randomized and nonrandomized participants to construct the measures. Posts without text (approximately 6% of the posts were excluded) and polls were excluded, which resulted in 761 posts and 9396 comments or replies across the 2 intervention arms.

The outcome of interest was on the postparticipant pair level; that is, whether each participant had engaged with (ie, commented or replied to) a post in the Facebook group (1 if yes and 0 if no). Comments are in response to the original post, whereas replies are responses to comments made by others on a post. We focused on comments and replies as these activities are active forms of engagement rather than passive types of engagement such as views and reactions (eg, "likes") and have been shown to positively predict weight loss [16,32,33]. We extracted engagement data from the private Facebook groups using the Grytics app [34]. The Grytics app allowed us to download all the content posted in each Facebook group into Microsoft Excel sheets along with its unique Facebook ID number (post ID, comment ID, and parent comment ID), time stamp, reaction data, and author. Participants were asked to allow the Grytics app to access their Facebook account name so that the content from the group was identified (ie, post or comment author name was included in export).

Post Characteristics

The post characteristics described in [Textbox 1](#) were included.

Textbox 1. Post characteristics.**Original poster**

- We used a binary variable to indicate whether the focal post was created by the interventionist (1) or participant (0).

Post length

- We measured the number of words in each post.

Content sentiment

- We measured the average sentiment (text polarity) of each post's content. Text polarity measures the valence and emotion in the text and ranges on a continuous spectrum from negative (lower value) to positive (higher value). We standardized the measure of sentiment for the analysis. Sentiment analysis was performed using the *sentimentr* package in R (version 3.6.1).

Topics

- We used natural language processing to identify the topics that appeared in each post, comment, and reply. The content was preprocessed to remove emojis and non-English characters. Topics were detected using *Top2Vec* in Python 3.10.0, a deep learning-based sentence embedding algorithm that detects topics in the documents. It detected 117 unique topics along with their top words (see [Multimedia Appendix 1](#) for a random sample of 35 topics with their top 8 words) from 10,157 pieces of content (posts, comments, and replies), which were further consolidated and coded under 7 topics based on the top 20 words of each original topic: exercise, diet, weight, MyFitnessPal app use, expressing emotion, sleep, and goals or plans. Here we focused on the topics of each post and created 7 binary variables to represent whether each post involves each of the aforementioned topics. Each post could include multiple topics; for example, a post mentioning a specific dietary goal would be categorized under both diet and goals or plans.

Time of the post

- We collected the time (number of days from day 1 of the intervention) and day of the week when the post was created.

Reply or Comment Characteristics

We constructed a series of variables representing the characteristics of replies or comments on each post. To reflect the content of conversations before each participant's engagement, if the participant engaged with the post, we

calculated these variables based on all previous comments or replies under the post before their engagement for each unique postparticipant pair; if the participant did not engage with the post, we calculated these measures based on all the comments or replies under the post. The characteristics are described in [Textbox 2](#).

Textbox 2. Reply or comment characteristics.**Tags or mentions**

- We created two binary variables to represent whether each participant had been tagged or mentioned by (1) interventionists or (2) other participants in the previous replies or comments within the same post. It is worth mentioning that most tags or mentions in our data were generated automatically by Facebook (eg, when participant A comments or replies to participant B's content, Facebook automatically generates a tag on B in A's reply or comment). Thus, most tags or mentions in our data represent reply or comment relationships. In very few instances, interventionists deliberately tagged previously disengaged participants; however, the sample size was too small to test their effects separately.

Reply or comment content sentiment

- We measured the average sentiment of all replies or comments for each postparticipant pair. The measure was standardized for the analysis.

Participants' Characteristics

The included participant characteristics were as follows:

- Percentage of previous posts commented or replied: For each post, we calculated the percentage of previous posts each participant has commented or replied to.
- Baseline and sociodemographic characteristics: Although these variables were not the focus of our analysis, we included baseline characteristics for each participant, including treatment condition (open vs closed), baseline weight, BMI, age, race, sex, education, marital status, number of people in the household, and employment status, as covariates in the analyses.

Statistical Analysis

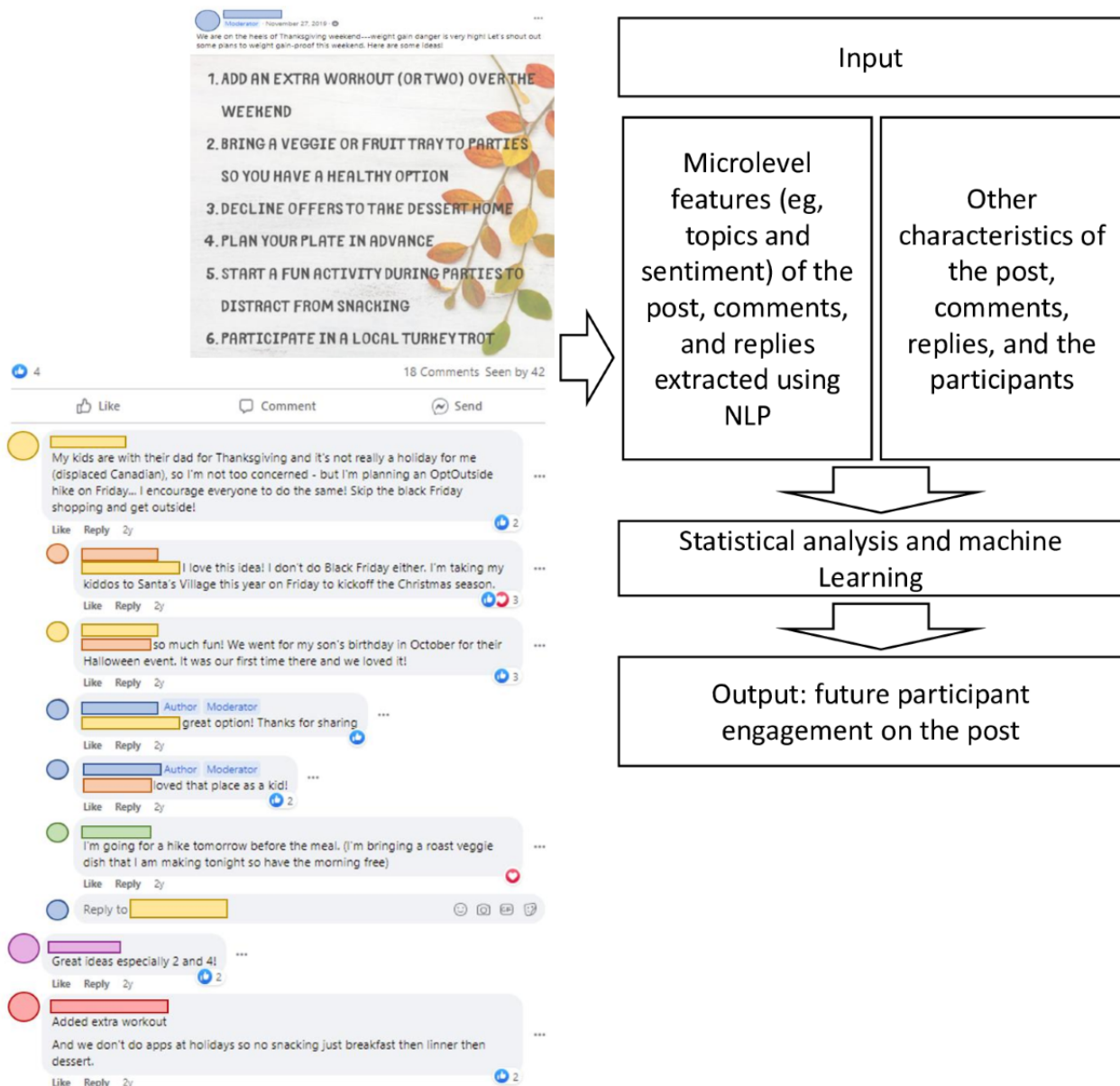
We focused our analysis on whether each randomized participant (N=80) had engaged with each post, as randomized participants had access to the Facebook group the entire length of the intervention (it should be noted that each post was only available in a particular treatment arm and, thus, can only be seen by 40 randomized participants). To examine what predicts participant engagement with each post, analyses were performed on the postparticipant pair level (ie, whether each participant engaged with each post). This allowed us to construct measures that accurately reflect the content (ie, posts and conversations) before each participant's engagement. We included all possible engagements (ie, instances where participants engaged and instances where they did not engage) from the 80 randomized

participants with each of the 761 posts, which resulted in a final sample of 31,968 observations (participants engaged in 4462 instances and did not engage in 27,506 instances) for our analysis.

The overall analysis framework is depicted in Figure 1. Data were screened for deviations from assumptions required for the used statistical analyses. We calculated descriptive statistics for the outcome and key independent variables for each treatment condition. To account for the fact that engagement was nested within each post, we performed a mixed-effects regression with postlevel random effects using participants' engagement as the primary outcome, along with all aforementioned key predictors

(ie, characteristics of the post, reply or comment, and the participants' previous engagement behavior) as independent variables, with participants' baseline and sociodemographic characteristics as covariates. We also included participant-level fixed effects as an alternative specification to account for possible omitted variable bias. As a robustness check, we also conducted a mixed-effects logistic regression with the same variables (Multimedia Appendix 1). To identify the important predictors of participant engagement, we reported the coefficient, 95% CI, and the associated P value for each predictor. All analyses were performed in STATA standard edition (version 16).

Figure 1. Analysis framework to identify important predictors of participant engagement. Left panel: an example of the intervention post and the comments or replies following it. Right panel: flow chart of the analysis. NLP: natural language processing.



Although regression analyses are useful to identify the statistical significance of linear relationships, some of the relationships might be much more complex (eg, nonlinear or moderated by other variables). To confirm the importance of significant predictors that we previously identified and to investigate how well these variables as a whole can predict participants'

engagement with a particular post, we included all aforementioned predictors in machine learning algorithms, including gradient boosting machines, deep learning models, and an ensemble of them [35,36], and examined the performance of these models by calculating key metrics, including (1) area under the curve (ranging from 0.5 to 1, with 1 being the best)

using a 5-fold cross-validation, (2) variable importance across different models, and (3) out-of-sample prediction accuracies on a balanced sample with 1600 observations (800 random observations with engagement and 800 observations without). In the machine learning model, variable importance was determined by calculating the relative influence of each variable: in the tree-based model, it was calculated based on whether that variable was selected or included during the tree-building process and how much it improves the model fit. In other non-tree-based models, it was calculated as the magnitude of the weight or coefficient connecting a specific input or variable to the output [37]. We also evaluated the variable importance using an alternative approach called Shapley Additive Explanations contribution in one of the top performing models, which measures how much the average model prediction would change with and without a specific feature or variable [38], as shown in [Multimedia Appendix 1](#). All the analyses were performed using the *h2o* package [39] in R (R Foundation for Statistical Computing; version 3.6.1).

Results

[Table 1](#) presents the sociodemographic characteristics and key measures of engagement for the participants (N=80). The mean age of participants was 40.2 (SD 11.2) years with a mean baseline BMI of 34.4 (SD 5.0) kg/m². Participants were predominantly female (68/80, 85%) and mostly White (72/80,

90%) White, with most (58/80, 73%) reporting completing college or graduate school. The baseline characteristics of the participants in the 2 treatment conditions were similar, and we did not observe significant differences in these variables between the groups. Over the 16-week intervention, participants in the closed group commented or replied to 9.9% (37/374; SD 10.4%) of the posts on average, whereas participants in the open group commented or replied to 8.8% (34/387; SD 9.7%) of the posts on average.

[Table 2](#) presents the key characteristics of the posts and comments or replies generated over the 16-week intervention across the 2 treatment arms. Post length was 33.4 words on average (SD 23.81) and 57.3% (436/761) of the posts were created by the interventionists. It should be noted participants in the 2 groups were exposed to identical program posts (whereas the number of self-generated posts by interventionists could be different). Topic modeling results showed that diet was the most popular topic across all posts (310/761, 40.7%), followed by exercise (163/761, 21.4%), goal or plan (152/761, 19.9%), and weight (138/761, 18.1%). We did not observe significant differences in post or comment or reply characteristics across groups, except that the percentage of replies or comments directed at randomized participants was significantly higher in the closed group than the open group ($P<.001$), possibly because there were 54 nonrandomized participants in the open group.

Table 1. Participant characteristics (N=80).

| Participant characteristics | Closed enrollment (n=40) | Open enrollment (n=40) |
|--|--------------------------|------------------------|
| Age (years), mean (SD) | 40.4 (11.8) | 40.0 (10.6) |
| Female, n (%) | 34 (85) | 34 (85) |
| Baseline BMI (kg/m ²), mean (SD) | 34.8 (5.4) | 34.0 (4.6) |
| Hispanic or Latino, n (%) | 3 (8) | 1 (3) |
| Race, n (%) | | |
| White | 36 (90) | 36 (90) |
| Black or African American | 3 (8) | 3 (8) |
| Asian | 0 (0) | 0 (0) |
| Native Hawaiian or other Pacific Islander | 0 (0) | 0 (0) |
| American Indian or Alaska Native | 0 (0) | 0 (0) |
| Multiethnic | 0 (0) | 1 (3) |
| Unknown | 1 (3) | 0 (0) |
| Marital status, n (%) | | |
| Married or living with partner but not married | 29 (73) | 30 (75) |
| Single | 8 (20) | 6 (15) |
| Widowed, divorced, or separated | 3 (8) | 4 (10) |
| Education, n (%) | | |
| Less than high school, high school degree, GED ^a , equivalent | 1 (3) | 2 (5) |
| Trade, technical, some college, associates | 8 (20) | 11 (28) |
| Bachelor's degree or some graduate school | 21 (53) | 17 (43) |
| Graduate degree | 10 (25) | 10 (25) |
| Employment status, n (%) | | |
| Employed full-time | 28 (70) | 27 (68) |
| Employed part-time | 7 (18) | 4 (10) |
| Student | 2 (5) | 2 (5) |
| Unemployed, retired, disabled, or homemaker | 3 (8) | 6 (15) |

^aGED: General Educational Development.

Table 2. Post and reply or comment characteristics over the 16-week intervention.

| Post characteristics | Closed enrollment (n=374) | Open enrollment (n=387) |
|---|---------------------------|-------------------------|
| Content sentiment, mean (SD) | 0.134 (0.197) | 0.133 (0.195) |
| Number of words, mean (SD) | 33.78 (24.63) | 33.28 (23.12) |
| Created by interventionists, n (%) | 225 (60.2) | 211 (54.5) |
| Topic, n (%) | | |
| Exercise | 80 (21.4) | 83 (21.4) |
| Diet | 158 (42.2) | 152 (39.3) |
| Weight | 64 (17.1) | 74 (19.1) |
| MyFitnessPal app | 61 (16.3) | 67 (17.3) |
| Expressing emotion | 28 (7.5) | 28 (7.2) |
| Sleep | 6 (1.6) | 5 (1.3) |
| Goals or plans | 80 (21.4) | 72 (18.6) |
| Reply or comment characteristics^a | | |
| Content sentiment, mean (SD) | 0.171 (0.255) | 0.156 (0.234) |
| Participant's reply to other participants, n (%) | 750 (23.8) | 803 (20.9) |
| Interventionist reply to a participant, n (%) | 1018 (32.3) | 1195 (30.9) |

^aClosed enrollment n=3152 and open enrollment n=6244.

Table 3 shows the results from mixed-effects regression models on how well each variable predicted participants' engagement with a specific post. Regarding post characteristics, we found that the overall probability of participants' engagement with each post decreased by 0.04% each day ($P < .001$; 95% CI 0.02%-0.06%). Participants had a 6.3% ($P < .001$; 95% CI 5.1%-7.5%) higher probability of engaging with posts generated by the interventionists than with posts created by other participants. Post length also mattered—one additional word in a post's content was associated with a 0.05% ($P < .001$; 95% CI 0.03%-0.08%) increase in participants' probability of engagement. Participants also had a 6.5% ($P < .001$; 95% CI 4.9%-8.1%) and 6.1% ($P < .001$; 95% CI 4.1%-8.1%) higher probability of engaging with posts if the post content was related to weight and goals or plans, respectively. Regarding reply or comment characteristics, participants were 44.8% ($P < .001$; 95% CI 42.8%-46.9%) more likely to engage with a post when they received replies from other participants in the conversation or 46% ($P < .001$; 95% CI 44.1%-48.0%) more likely to engage if they received replies by interventionists. A 1 SD decrease in the sentiment in the previous replies or comments was associated with a 5.4% ($P < .001$; 95% CI 4.9%-5.9%) increase in the probability of participant engagement. Participants' engagement in previous posts was a strong predictor of future engagement—a 1% increase in participants' previous engagement was associated with a 0.76% ($P < .001$; 95% CI 0.74%-0.79%) increase in their probability to engage with a subsequent post. Robustness analyses showed that these results were largely consistent with the results from (1) multivariate linear regression results and

mixed-effects regression with participant-level fixed effects and (2) mixed-effects logistic regression with or without postlevel random effects and participant fixed effects. Details of the results from these additional regression analyses can be found in [Multimedia Appendix 1](#).

To confirm the importance of previously identified predictors and test how well the aforementioned variables can predict the probability of a participant engaging with a post, we performed a variety of machine learning-based classification algorithms with all the aforementioned predictors as the input and participant engagement as the outcome. Of the 32 models we tested, the ensemble approach of gradient boosting machine learning-based and deep learning-based classification algorithms performed the best, with an average area under the curve of 0.963 using 5-fold cross-validation (see more results in [Multimedia Appendix 1](#)). **Figure 2** shows the variable importance across 20 machine learning models (excluding ensemble models) we tested, with those indicated by yellow and red being more important variables. The results show that receiving a reply from other participants and interventionists, percentage of previous posts participants had engaged in, average sentiment in previous replies or comments, time of the post, and day of the week were the most important variables across models, which were consistent with the results from regression analyses. Finally, we performed out-of-sample predictions on a balanced sample with 1600 observations (800 observations with engagement and 800 observations without) and achieved 90.9% accuracy and 0.908 F_1 score at maximum.

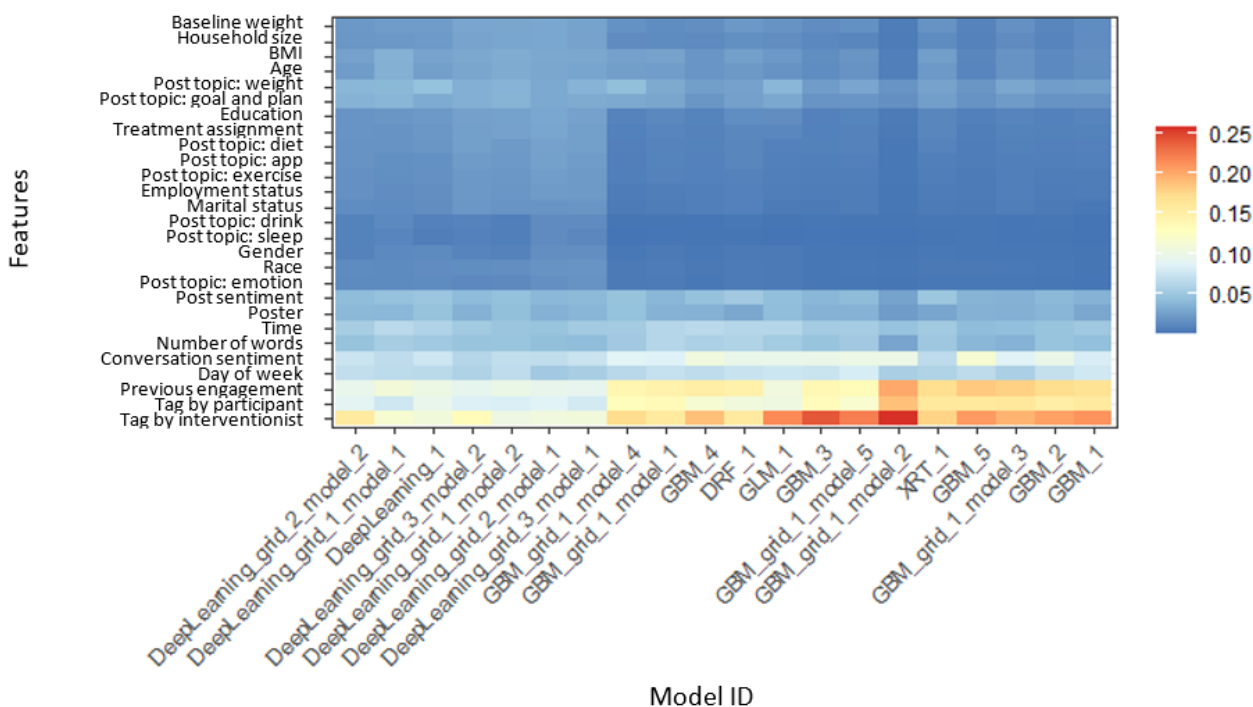
Table 3. Mixed-effects regression results predicting participants' engagement (N=31,968)^a.

| Mixed-effects regression | Values, mean (SD; range) | Coefficient (95% CI) | P value |
|--|--------------------------|------------------------------|---------|
| Outcome: participants' engagement | 0.140 (0.347; 0 to 1) | — ^b | — |
| Post characteristics | | | |
| Created by interventionists | 0.583 (0.493; 0 to 1) | 0.0627 (0.0507 to 0.0746) | <.001 |
| Number of words | 33.44 (23.80; 1 to 107) | 0.0005 (0.0003 to 0.0008) | <.001 |
| Content sentiment (standardized) | 0 (1; -3.67 to 4.74) | -.0042 (-0.097 to 0.0012) | .13 |
| Topics | | | |
| Exercise | 0.212 (0.409; 0 to 1) | 0.0096 (-0.0075 to 0.0266) | .27 |
| Diet | 0.389 (0.487; 0 to 1) | -0.0085 (-0.0249 to 0.0078) | .31 |
| Weight | 0.191 (0.392; 0 to 1) | 0.0654 (0.0494 to 0.0814) | <.001 |
| MyFitnessPal app | 0.163 (0.369; 0 to 1) | -0.0377 (-0.0534 to -0.0219) | <.001 |
| Expressing emotion | 0.071 (0.257; 0 to 1) | 0.0083 (-0.01558 to 0.0321) | .50 |
| Sleep | 0.0141 (0.117; 0 to 1) | -0.0587 (-0.1070 to -0.0103) | .02 |
| Goals or plans | 0.209 (0.407; 0 to 1) | 0.0612 (0.0414 to 0.0811) | <.001 |
| Time of the post | 46.61 (32.94; 1 to 112) | -0.0004 (-0.0006 to -0.0002) | <.001 |
| Reply or comment characteristics | | | |
| Content sentiment (standardized) | 0 (1; -7.84 to 5.91) | -0.0539 (-0.0589 to -0.0488) | <.001 |
| Replied by other participants | 0.026 (0.161; 0 to 1) | 0.4484 (0.4279 to 0.4690) | <.001 |
| Replied by interventionists | 0.029 (0.167; 0 to 1) | 0.4604 (0.4409 to 0.4798) | <.001 |
| Participants characteristics | | | |
| Percentage previous posts commented or replied | 13.16 (13.68; 0 to 100) | 0.0076 (0.0074 to 0.0079) | <.001 |

^aThe model included postlevel random effects and also controlled for day of the week when the post was created and other baseline and sociodemographic characteristics of the participants, including treatment assignment, race, marital status, education, employment, number of people in the household, age, gender, baseline BMI, and weight.

^bNot available.

Figure 2. Variable importance of predicting participant engagement across 20 machine learning models. The x-axis shows different model names, and variables from top to bottom on the y-axis are baseline weight, number of people in the household, baseline BMI, age, post topic weight, post topic goal and plan, education, treatment assignment, post topic diet, post topic app, post topic exercise, employment status, marital status, post topic drink, post topic sleep, gender, race, post topic expressing emotion, post sentiment, whether the post is created by interventionists, day of the intervention when the post is created, word count of the post, replies or comments sentiment, day of the week when the post is created, percentage of previous posts engaged, and whether replied by other participants or by interventionists.



Discussion

Principal Findings

In this study, we conducted secondary analyses using data from a 2-arm pilot feasibility randomized controlled trial that delivered lifestyle interventions via Facebook. We analyzed commenting or replying behavior from 80 participants in response to each of the 761 posts generated by counselors and participants over the 16-week intervention period and linked them to predictors, including post characteristics (eg, time, post length, and post topic), conversation characteristics (eg, sentiment of the conversation and participants being replied to), and participant characteristics (eg, sociodemographics and previous commenting or replying behavior). Our findings suggest that although participants’ comments or replies decreased over time, important characteristics of the post, the conversation attached to that post, and the participants’ engagement patterns predicted whether a participant engaged with a specific post. For example, we found that participants who engaged more with prior posts were more likely to engage with future posts. Posts that were longer (with the maximum number of words not exceeding 107), were created by interventionists, or had content related to weight (eg, weigh-in posts) and goal setting are more likely to attract engagement. The latter is consistent with the design of the intervention—participants were asked to set diet and exercise goals for the week each Monday, report their progress on their goals on Sunday, and report their weight change for the week each Friday. This is also encouraging because goal setting [40] and frequent self-weighing [41] are key behavioral weight loss

Model ID

strategies. Furthermore, posts with replies or comments that directly mentioned the focal participant were much more likely to attract subsequent replies from that participant. Moreover, posts with replies or comments that contained negative sentiments were more likely to attract subsequent comments. This is likely because participants who share struggles, problems, and challenges are naturally more negative in sentiment (see [Multimedia Appendix 1](#) for examples of replies or comments with negative sentiments), and such content often attracts support and brainstorming from other participants who may feel called to help when others are struggling. These results were robust to multiple alternative specifications. Machine learning results also show that together, these characteristics can predict participant engagement with a high accuracy of 90.9%.

Implications

In this study, we demonstrate the potential of using NLP tools to analyze microlevel conversational data and identify factors influencing participants’ commenting or replying behavior in a social media–delivered weight loss intervention. Our findings shed light on some important microlevel characteristics of the participants, posts, and conversations, which can shape participants’ experiences during the intervention and predict their future engagement. These results have implications for the design and implementation of social media–delivered behavioral interventions in ways that maximize participant engagement. We previously reported a strong association between participant engagement and weight loss [31], which suggests that engagement-enhancing strategies could lead to more favorable outcomes. For example, enhancing engagement

early on may help with sustained engagement. Furthermore, we found that receiving replies appears to stimulate further engagement from the participant. This may also be a function of whether a participant shares something substantive about themselves in a comment. For example, if a participant's comment is "thanks!" the interventionist may just hit the "like" button; however, if a participant's comment is a question or sharing of a goal, the interventionist and other participants are more likely to reply to continue that conversation. Further research should explore the type of comments (eg, questions, sharing a problem, and setting a goal) that are most likely to elicit replies from others. Program content should be designed in ways that nudge participants to post more often and interact more with each other (eg, use open-ended questions and encourage peer-to-peer support). Although in this study we found that participants were more likely to respond to interventionists, greater peer-to-peer engagement could also facilitate stronger group cohesion, thereby further enhancing participant engagement. Posts in which participants share struggles, problems, and challenges they have encountered during the weight loss process may draw more participants into the conversation, which may generate richer brainstorming and social support, both of which could also enhance group cohesion. These implications could be applicable not only to social media-delivered weight loss interventions but also to other digital health interventions more generally.

Comparison With Prior Work

Although many studies have tested social media-delivered weight loss interventions or emphasized the importance of participant engagement in web-based communities [12-15,42], only a handful of studies have identified the factors that can influence participant engagement in digital health interventions [43,44], and most of them focused on macrolevel factors such as post type and participants' characteristics (eg, age and gender) [9,14]. Few studies have examined microlevel factors such as conversation dynamics. Several studies recognized this limitation and called for more research to identify all relevant factors that can predict participant engagement [28,43,45,46]. This study contributes to this line of research in two ways: (1) we demonstrate how content generated by interventionists and participants during the course of a digital intervention can be leveraged and combined with data science and NLP tools to identify microlevel predictors of participant engagement, and (2) we have identified many microlevel factors that influence participant engagement, which, to the best of our knowledge, have not been studied in previous social media-based behavioral interventions. This has practical implications for future intervention designs that can maximize participant engagement.

Similar to previous studies, we found that participant engagement is highly variable [17-19], and it generally declines over time [16,47]. Although many factors identified in this study have not been studied in the social media-based behavioral intervention context, our findings are consistent with psychological and sociological theories, as well as several empirical research on web-based communication. For example, our finding that posts created by interventionists are more popular is consistent with other studies on web-based communities, showing that important users or those with

designated roles are more likely to draw responses from other users [26], possibly because of preferential attachment [48]. Our finding that replies or comments with negative sentiments draw more engagement implies that participants are more likely to reply or comment when they see others sharing their struggles and challenges. This could be possibly explained by social support processes, which have substantial empirical support across various web-based settings [22-25,49]. Finally, the importance of being replied to by interventionists and other participants can be explained by the preference for reciprocity [50], which has been found to be an important driver for communication in many other web-based settings [20,21].

Limitations and Future Work

This study has several limitations that point to avenues for future research. First, our sample size was small (80 participants; 10,157 total posts, comments, and replies) and our participants were predominantly White (72/80, 90%) and female (68/80, 85%). This limits the generalizability of our results, following a long-standing pattern in weight loss studies of difficulty recruiting male participants [51]. Similarly, given that 96% (77/80) of our participants reported attaining a college degree or advanced degree, we cannot generalize our results to individuals with lower levels of education. Future studies should devise recruitment strategies that attract more male participants and participants with low levels of education to further explore the individual heterogeneity across people from different backgrounds. Second, this study did not fully tease out all possible confounding factors and thus cannot establish causality. For instance, participants who are more successful in losing weight might also be more likely to comment simply because they are paying more attention to the group and have more to say as they are applying the knowledge and strategies they are learning. Future studies should include larger trials, surveys with more longitudinal measures (eg, physical activity and diet tracking, mental health, and other behaviors), and qualitative studies to establish the possible bidirectional and causal relationships between engagement and these factors. Third, we focused on replies and comments and did not explore other types of engagement such as reactions and views of posts and comments. Although comments and replies have been considered more substantive than other engagements (eg, likes and "lurking"), other engagements potentially comprise a substantial proportion of social media use and thus warrant careful consideration in future studies [33]. Similarly, although we included participants' posts in our analysis, we did not include posts only with images or videos or investigate the factors influencing participants' decisions to create posts, which is another important form of engagement. In addition, although we included tags or mentions relationships in our study, most tags were automatically generated by Facebook during replies or comments. Future studies should consider whether deliberate tagging can nudge disengaged participants to re-engage with the program. Fourth, we assumed all participants had an opportunity to engage with every post in the group, and we considered all replies or comments when constructing the predictors for participants who did not engage with a certain post, which might not necessarily be the case if the participant did not see the post or the previous comments or replies. Future

studies could take additional information into account, such as what participants viewed and the time a participant spends on Facebook. This will allow researchers to construct more refined measures to reflect the condition under which participants make a decision of whether to engage. Finally, although we observed that longer posts are more likely to draw engagement from participants, it should be noted the posts in this intervention were generally short by design (mean 33.4, maximum 107 words). Social media marketing reports reveal that Facebook posts that have <50 characters receive the highest level of engagement relative to longer posts [52]. A/B testing of a wide range of post lengths and different types of posts (eg, goal setting vs problem sharing) is needed to determine the ideal length of posts to maximize engagement in behavioral interventions.

Conclusions

In this study, we performed secondary analyses using data from a pilot feasibility randomized weight loss trial that delivered a lifestyle intervention via Facebook and linked participants' engagement with several important predictors, including characteristics of the posts, replies or comments, and participants. Our results point to the potential of using data science and NLP tools to analyze microlevel behavioral or conversational data and identify factors influencing participants' engagement during the social media weight loss intervention, which have implications for the design and implementation of future interventions that could lead to more favorable weight loss outcomes. Future studies are warranted to validate our results and further explore these relationships in similar and larger trials.

Acknowledgments

This project was supported by the National Institutes of Health grant K24HL124366 (principal investigator: SP).

Conflicts of Interest

SP has been a paid advisor for WW (formerly Weight Watchers) and FitBit.

Multimedia Appendix 1

Supplementary analyses and results.

[\[DOCX File, 91 KB-Multimedia Appendix 1\]](#)

References

1. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity and severe obesity among adults: United States, 2017-2018. NCHS Data Brief 2020 Feb(360):1-8 [FREE Full text] [Medline: 32487284]
2. Wolin KY, Carson K, Colditz GA. Obesity and cancer. *Oncologist* 2010;15(6):556-565 [FREE Full text] [doi: 10.1634/theoncologist.2009-0285] [Medline: 20507889]
3. Pagoto S. The current state of lifestyle intervention implementation research: where do we go next? *Transl Behav Med* 2011 Sep;1(3):401-405 [FREE Full text] [doi: 10.1007/s13142-011-0071-x] [Medline: 24073065]
4. Beleigoli AM, Andrade AQ, Caçado AG, Paulo MN, Diniz MD, Ribeiro AL. Web-based digital health interventions for weight loss and lifestyle habit changes in overweight and obese adults: systematic review and meta-analysis. *J Med Internet Res* 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
5. Pagoto S, Waring ME, May CN, Ding EY, Kunz WH, Hayes R, et al. Adapting behavioral interventions for social media delivery. *J Med Internet Res* 2016 Jan 29;18(1):e24 [FREE Full text] [doi: 10.2196/jmir.5086] [Medline: 26825969]
6. Waring ME, Jake-Schoffman DE, Holovatska MM, Mejia C, Williams JC, Pagoto SL. Social media and obesity in adults: a review of recent research and future directions. *Curr Diab Rep* 2018 Apr 18;18(6):34. [doi: 10.1007/s11892-018-1001-9] [Medline: 29671135]
7. Fox S. The social life of health information. Pew Research Center. URL: <https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/> [accessed 2022-06-22]
8. Heo Y, Park J, Kim J, Park H. The emerging viewertariat in South Korea: The Seoul mayoral TV debate on Twitter, Facebook, and blogs. *Telematic Inform* 2016 May;33(2):570-583. [doi: 10.1016/j.tele.2015.08.003]
9. Rus HM, Cameron LD. Health communication in social media: message features predicting user engagement on diabetes-related facebook pages. *Ann Behav Med* 2016 Oct;50(5):678-689. [doi: 10.1007/s12160-016-9793-9] [Medline: 27059761]
10. Nam Y, Lee Y, Park HW. Measuring web ecology by Facebook, Twitter, blogs and online news: 2012 general election in South Korea. *Qual Quant* 2014 Mar 16;49(2):675-689. [doi: 10.1007/s11135-014-0016-9]
11. An R, Ji M, Zhang S. Effectiveness of social media-based interventions on weight-related behaviors and body weight status: review and meta-analysis. *Am J Health Behav* 2017 Nov 01;41(6):670-682. [doi: 10.5993/AJHB.41.6.1] [Medline: 29025495]
12. Pagoto SL, Waring ME, Schneider KL, Oleski JL, Olenzki E, Hayes RB, et al. Twitter-delivered behavioral weight-loss interventions: a pilot series. *JMIR Res Protoc* 2015 Oct 23;4(4):e123 [FREE Full text] [doi: 10.2196/resprot.4864] [Medline: 26500186]

13. Turner-McGrievy GM, Tate DF. Weight loss social support in 140 characters or less: use of an online social network in a remotely delivered weight loss intervention. *Transl Behav Med* 2013 Sep;3(3):287-294 [FREE Full text] [doi: [10.1007/s13142-012-0183-y](https://doi.org/10.1007/s13142-012-0183-y)] [Medline: [24073180](https://pubmed.ncbi.nlm.nih.gov/24073180/)]
14. Hales SB, Davidson C, Turner-McGrievy GM. Varying social media post types differentially impacts engagement in a behavioral weight loss intervention. *Transl Behav Med* 2014 Dec;4(4):355-362 [FREE Full text] [doi: [10.1007/s13142-014-0274-z](https://doi.org/10.1007/s13142-014-0274-z)] [Medline: [25584084](https://pubmed.ncbi.nlm.nih.gov/25584084/)]
15. Pagoto S, Waring M, Olendzki E, Oleski J, May C, Evans M. The feasibility of incentivizing participation in an online social network weight loss program. In: *Proceedings of the Hawaii International Conference on System Sciences*. 2017 Presented at: Hawaii International Conference on System Sciences; Jan 4-7, 2017; Hawaii, USA.
16. Turner-McGrievy G, Tate D. Tweets, apps, and pods: results of the 6-month mobile pounds off digitally (mobile POD) randomized weight-loss intervention among adults. *J Med Internet Res* 2011 Dec 20;13(4):e120 [FREE Full text] [doi: [10.2196/jmir.1841](https://doi.org/10.2196/jmir.1841)] [Medline: [22186428](https://pubmed.ncbi.nlm.nih.gov/22186428/)]
17. Pagoto S, Waring M, Jake-Schoffman D. What type of engagement predicts success in a Facebook weight loss group? In: *Proceedings of the Hawaii International Conference on System Sciences*. 2018 Presented at: Hawaii International Conference on System Sciences; Jan 3 – 6, 2018; Hawaii, USA. [doi: [10.24251/hicss.2018.419](https://doi.org/10.24251/hicss.2018.419)]
18. West DS, Monroe CM, Turner-McGrievy G, Sundstrom B, Larsen C, Magrader K, et al. A technology-mediated behavioral weight gain prevention intervention for college students: controlled, quasi-experimental study. *J Med Internet Res* 2016 Jun 13;18(6):e133 [FREE Full text] [doi: [10.2196/jmir.5474](https://doi.org/10.2196/jmir.5474)] [Medline: [27296086](https://pubmed.ncbi.nlm.nih.gov/27296086/)]
19. Herring SJ, Cruice JF, Bennett GG, Rose MZ, Davey A, Foster GD. Preventing excessive gestational weight gain among African American women: a randomized clinical trial. *Obesity (Silver Spring)* 2016 Jan;24(1):30-36 [FREE Full text] [doi: [10.1002/oby.21240](https://doi.org/10.1002/oby.21240)] [Medline: [26592857](https://pubmed.ncbi.nlm.nih.gov/26592857/)]
20. Constant D, Sproull L, Kiesler S. The kindness of strangers: the usefulness of electronic weak ties for technical advice. *Org Sci* 1996 Apr;7(2):119-135. [doi: [10.1287/orsc.7.2.119](https://doi.org/10.1287/orsc.7.2.119)]
21. How open source software works: “free” user-to-user assistance. In: *Produktentwicklung mit virtuellen Communities*. Wiesbaden, Germany: Springer Verlag; 2004.
22. Kramer AD, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci U S A* 2014 Jun 17;111(24):8788-8790 [FREE Full text] [doi: [10.1073/pnas.1320040111](https://doi.org/10.1073/pnas.1320040111)] [Medline: [24889601](https://pubmed.ncbi.nlm.nih.gov/24889601/)]
23. Kramer A. The spread of emotion via Facebook. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2012 Presented at: CHI '12: CHI Conference on Human Factors in Computing Systems; May 5 - 10, 2012; Austin Texas USA. [doi: [10.1145/2207676.2207787](https://doi.org/10.1145/2207676.2207787)]
24. Song Y, Lin Q, Kwon KH, Choy CH, Xu R. Contagion of offensive speech online: an interactional analysis of political swearing. *Comput Human Behav* 2022 Feb;127:107046. [doi: [10.1016/j.chb.2021.107046](https://doi.org/10.1016/j.chb.2021.107046)]
25. Centola D. The spread of behavior in an online social network experiment. *Science* 2010 Sep 03;329(5996):1194-1197. [doi: [10.1126/science.1185231](https://doi.org/10.1126/science.1185231)] [Medline: [20813952](https://pubmed.ncbi.nlm.nih.gov/20813952/)]
26. Johnson SL, Faraj S, Kudaravalli S. Emergence of power laws in online communities: the role of social mechanisms and preferential attachment. *MIS Q* 2014 3;38(3):795-808. [doi: [10.25300/misq/2014/38.3.08](https://doi.org/10.25300/misq/2014/38.3.08)]
27. Li J, Zheng H, Duan X. Factors influencing the popularity of a health-related answer on a Chinese question-and-answer website: case study. *J Med Internet Res* 2021 Sep 28;23(9):e29885 [FREE Full text] [doi: [10.2196/29885](https://doi.org/10.2196/29885)] [Medline: [34581675](https://pubmed.ncbi.nlm.nih.gov/34581675/)]
28. Pagoto S, Waring ME. A call for a science of engagement: comment on rus and cameron. *Ann Behav Med* 2016 Oct;50(5):690-691. [doi: [10.1007/s12160-016-9839-z](https://doi.org/10.1007/s12160-016-9839-z)] [Medline: [27663577](https://pubmed.ncbi.nlm.nih.gov/27663577/)]
29. Guetterman TC, Chang T, DeJonckheere M, Basu T, Scruggs E, Vydiswaran VG. Augmenting qualitative text analysis with natural language processing: methodological study. *J Med Internet Res* 2018 Jun 29;20(6):e231 [FREE Full text] [doi: [10.2196/jmir.9702](https://doi.org/10.2196/jmir.9702)] [Medline: [29959110](https://pubmed.ncbi.nlm.nih.gov/29959110/)]
30. Wang ML, Waring ME, Jake-Schoffman DE, Oleski JL, Michaels Z, Goetz JM, et al. Clinic versus online social network-delivered lifestyle interventions: protocol for the get social noninferiority randomized controlled trial. *JMIR Res Protoc* 2017 Dec 11;6(12):e243 [FREE Full text] [doi: [10.2196/resprot.8068](https://doi.org/10.2196/resprot.8068)] [Medline: [29229591](https://pubmed.ncbi.nlm.nih.gov/29229591/)]
31. Pagoto SL, Schroeder MW, Xu R, Waring ME, Groshon L, Goetz JM, et al. A Facebook-delivered weight loss intervention using open enrollment: randomized pilot feasibility trial. *JMIR Form Res* 2022 May 06;6(5):e33663 [FREE Full text] [doi: [10.2196/33663](https://doi.org/10.2196/33663)] [Medline: [35522466](https://pubmed.ncbi.nlm.nih.gov/35522466/)]
32. Cavallo DN, Martinez R, Webb Hooper M, Flocke S. Feasibility of a social media-based weight loss intervention designed for low-SES adults. *Transl Behav Med* 2021 Apr 26;11(4):981-992 [FREE Full text] [doi: [10.1093/tbm/ibaa070](https://doi.org/10.1093/tbm/ibaa070)] [Medline: [32716040](https://pubmed.ncbi.nlm.nih.gov/32716040/)]
33. Edelman N. What is lurking? A literature review of research on lurking. In: *The Psychology of Social Networking Vol.1*. Warsaw, Poland: De Gruyter Open Poland; 2016.
34. Grytics. URL: <https://grytics.com/> [accessed 2022-07-13]
35. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neurobot* 2013;7:21 [FREE Full text] [doi: [10.3389/fnbot.2013.00021](https://doi.org/10.3389/fnbot.2013.00021)] [Medline: [24409142](https://pubmed.ncbi.nlm.nih.gov/24409142/)]

36. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, Massachusetts, United States: MIT Press; 2016.
37. Rifkin R, Klautau A. In defense of one-vs-all classification. *J Mach Learn Res* 2004;5:101-141. [doi: [10.5555/1005332.1005336](https://doi.org/10.5555/1005332.1005336)]
38. Lundberg S, Lee S. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4 - 9, 2017; Long Beach California USA.
39. R Interface for 'H2O'. Package 'h2o'. 2018. URL: <https://cran.microsoft.com/snapshot/2018-10-04/web/packages/h2o/h2o.pdf> [accessed 2022-07-11]
40. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. *BMC Public Health* 2011 Feb 18;11:119 [FREE Full text] [doi: [10.1186/1471-2458-11-119](https://doi.org/10.1186/1471-2458-11-119)] [Medline: [21333011](https://pubmed.ncbi.nlm.nih.gov/21333011/)]
41. Steinberg DM, Bennett GG, Askew S, Tate DF. Weighing every day matters: daily weighing improves weight loss and adoption of weight control behaviors. *J Acad Nutr Diet* 2015 Apr;115(4):511-518 [FREE Full text] [doi: [10.1016/j.jand.2014.12.011](https://doi.org/10.1016/j.jand.2014.12.011)] [Medline: [25683820](https://pubmed.ncbi.nlm.nih.gov/25683820/)]
42. Cavallo DN, Sisneros JA, Ronay AA, Robbins CL, Jilcott Pitts SB, Keyserling TC, et al. Assessing the feasibility of a web-based weight loss intervention for low-income women of reproductive age: a pilot study. *JMIR Res Protoc* 2016 Feb 26;5(1):e30 [FREE Full text] [doi: [10.2196/resprot.4865](https://doi.org/10.2196/resprot.4865)] [Medline: [26920252](https://pubmed.ncbi.nlm.nih.gov/26920252/)]
43. Escobar-Viera CG, Melcher EM, Miller RS, Whitfield DL, Jacobson-López D, Gordon JD, et al. A systematic review of the engagement with social media-delivered interventions for improving health outcomes among sexual and gender minorities. *Internet Interv* 2021 Sep;25:100428 [FREE Full text] [doi: [10.1016/j.invent.2021.100428](https://doi.org/10.1016/j.invent.2021.100428)] [Medline: [34401387](https://pubmed.ncbi.nlm.nih.gov/34401387/)]
44. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. *Int J Behav Nutr Phys Act* 2016 Dec 07;13(1):127 [FREE Full text] [doi: [10.1186/s12966-016-0454-y](https://doi.org/10.1186/s12966-016-0454-y)] [Medline: [27927218](https://pubmed.ncbi.nlm.nih.gov/27927218/)]
45. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. *J Am Med Inform Assoc* 2015 Jan;22(1):243-256 [FREE Full text] [doi: [10.1136/amiainl-2014-002841](https://doi.org/10.1136/amiainl-2014-002841)] [Medline: [25005606](https://pubmed.ncbi.nlm.nih.gov/25005606/)]
46. Hale TM, Pathipati AS, Zan S, Jethwani K. Representation of health conditions on Facebook: content analysis and evaluation of user engagement. *J Med Internet Res* 2014 Aug 04;16(8):e182 [FREE Full text] [doi: [10.2196/jmir.3275](https://doi.org/10.2196/jmir.3275)] [Medline: [25092386](https://pubmed.ncbi.nlm.nih.gov/25092386/)]
47. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelandotte C. Are health behavior change interventions that use online social networks effective? A systematic review. *J Med Internet Res* 2014 Feb 14;16(2):e40 [FREE Full text] [doi: [10.2196/jmir.2952](https://doi.org/10.2196/jmir.2952)] [Medline: [24550083](https://pubmed.ncbi.nlm.nih.gov/24550083/)]
48. Barabasi AL, Albert R. Emergence of scaling in random networks. *Science* 1999 Oct 15;286(5439):509-512. [doi: [10.1126/science.286.5439.509](https://doi.org/10.1126/science.286.5439.509)] [Medline: [10521342](https://pubmed.ncbi.nlm.nih.gov/10521342/)]
49. White M, Dorman SM. Receiving social support online: implications for health education. *Health Educ Res* 2001 Dec;16(6):693-707. [doi: [10.1093/her/16.6.693](https://doi.org/10.1093/her/16.6.693)] [Medline: [11780708](https://pubmed.ncbi.nlm.nih.gov/11780708/)]
50. Exchange and Power in Social Life. Milton Park, Abingdon-on-Thames, Oxfordshire United Kingdom: Taylor & Francis; 1964.
51. Franz MJ, VanWormer JJ, Crain AL, Boucher JL, Histon T, Caplan W, et al. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. *J Am Diet Assoc* 2007 Oct;107(10):1755-1767. [doi: [10.1016/j.jada.2007.07.017](https://doi.org/10.1016/j.jada.2007.07.017)] [Medline: [17904936](https://pubmed.ncbi.nlm.nih.gov/17904936/)]
52. Zote J. How long should social posts be? Try this social media character counter. SproutSocial. URL: <https://sproutsocial.com/insights/social-media-character-counter/> [accessed 2022-06-22]

Abbreviations

DPP: Diabetes Prevention Program

NLP: natural language processing

Edited by A Mavragani; submitted 17.03.22; peer-reviewed by H Alhoori, N Maglaveras, P Kaur; comments to author 23.05.22; revised version received 22.06.22; accepted 23.06.22; published 28.07.22

Please cite as:

Xu R, Divito J, Bannor R, Schroeder M, Pagoto S

Predicting Participant Engagement in a Social Media-Delivered Lifestyle Intervention Using Microlevel Conversational Data: Secondary Analysis of Data From a Pilot Randomized Controlled Trial

JMIR Form Res 2022;6(7):e38068

URL: <https://formative.jmir.org/2022/7/e38068>

doi: [10.2196/38068](https://doi.org/10.2196/38068)

PMID:

©Ran Xu, Joseph Divito, Richard Bannor, Matthew Schroeder, Sherry Pagoto. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 28.07.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.