

## Electronic Supplementary Material 1

### *Data acquisition and cleaning procedures:*

Data were collected from Twitter's official, public streaming Application Programming Interface (API), using Python 3.6 and the Python library Tweepy. The API was used to collect tweets that met certain keyword filters. Messages needed to have a least one of the following keywords or hashtags: 'talkmh', 'mentalhealthawareness', 'suicideawareness', 'suicideprevention', 'selfharmawareness', 'keeptalkingmh', 'bpdfam', 'sicknotweak', 'mentalhealthmatters', 'beatdepression', 'mentalhealth', and 'endthestigma'. These terms were identified using an iterative qualitative procedure that involved manually reviewing positive mental health messages and identifying additional related terms through examining co-occurrence of other terms/phrases as well as manually examining new, related terms automatically suggested by Twitter's search feature. Data were collected in real-time over August 7<sup>th</sup>, 2017, to August 14<sup>th</sup>, 2017. Qualitative examination of 100 randomly selected messages from this sample indicated that 98% of tweets captured were indeed the positive messages sought for this research. The initial sample of tweets were cleaned to remove retweets, duplicates and non-English messages, such that the sample only included originally authored messages. Data management and analyses were conducted in Python 3.6 and R 3.4.3.

### *Creation of predictor variables:*

Several types of predictor variables were created and examined. Some account characteristics (number of followers, number of prior posts, and media included) were present in meta-data automatically provided for each tweet by the Twitter API. The inclusion of a url link or communication directed at another user (requires the use of the @ symbol) were identified through regular-expression searches of each tweet's text. The presence of an emoji was identified using Python's emoji library, which is based on searching for all emoji unicodes in the text. The number of hashtags were identified using regular expression searches. Counting the various parts of speech in each tweet was performed automatically by using Python's Natural Language Toolkit (NLTK) library. Lastly, the percent of messages in each of the lexical categories were automatically calculated using the Empath library in Python [Fast, E., Chen, B., & Bernstein, M. S. (2016). Empath: Understanding topic signals in large-scale text. Paper presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems]. In brief, Empath is a lexicon, meaning that each category is represented by dozens of related words that define or characterize the particular category. For example, the "military" category includes words such as "patriot," "marine," and "soldier." Empath is an advanced lexicon recently developed by Stanford University and, as detailed in the publication by Fast et al., the lexicon for each word was created by using a neural network to analyze 1.8 billion words from literature in the English language, producing numerical representations of each word. These numerical representations allow for calculating the distance between each word and thereby determining how closely related two words are to each other. Each lexical category was created by examining automated clustering of words produced by the neural network as well as using subsequent human review and validation of each category. For the purposes of this manuscript, a tweet was coded into a certain lexical category if it had at least one keyword from that category's lexicon. The lexical categories we chose to examine for this research were 5 sentiment-related categories (achievement, positive emotion, negative emotion, sadness, and anger) for which the Empath lexicon has undergone validation and 5 topic categories (military, school, science, eating, and

appearance) which are categories of interest (e.g., veteran suicide) in the mental health literature and community of prevention practitioners.

*Supplementary Table: Example words representing each lexical category significantly associated with shared messages*

| <b>Lexical Category</b> | <b>Example Words</b>  |
|-------------------------|---|
| Achievement             | success, celebrate, goal, victory, praise, approval, raise, cheer, determination, applaud         |
| Sadness                 | unhappy, heartache, troubled, failure, disappoint, abandonment, loss, depressive, tragedy, suffer |
| Military                | uniform, battle, patriotic, deploy, army, recruiting, troop, captain, platoon, defender           |
| Eating                  | nutritious, food, plate, diet, bread, craving, meal, tasty, vegetarian, dessert                   |
| Appearance              | gorgeous, hairstyle, curvy, skinny, fashionable, slimming, physique, supermodel, tan, dyed        |