Research and Applications

Mapping gender transition sentiment patterns via social media data: toward decreasing transgender mental health disparities

Oliver L. Haimson

School of Information, University of Michigan, Ann Arbor, Michigan, USA

Corresponding Author: Oliver L. Haimson, PhD, MSc, School of Information, University of Michigan, 105 South State Street, Ann Arbor, MI 48109-1285, USA (haimson@umich.edu)

ABSTRACT

Objective: Transgender people face substantial mental health disparities, and this population's emotional well-being can be particularly volatile during gender transition. Understanding gender transition sentiment patterns can positively impact transgender people by enabling them to anticipate, and put support in place for, particularly difficult time periods. Yet, tracking sentiment over time throughout gender transition is challenging using traditional research methods. This study's objective was to use social media data to understand average gender transition sentiment patterns.

Materials and Methods: Computational sentiment analysis and statistics were used to analyze 41 066 posts from 240 Tumblr transition blogs (online spaces where transgender people document gender transitions) to understand sentiment patterns over time and quantify relationships between transgender identity disclosures, sentiment, and social support.

Results: Findings suggest that sentiment increases over time on average throughout gender transition, particularly when people receive supportive responses to transgender identity disclosures. However, after disclosures to family members, people experienced temporary increased negative sentiment, followed by increased positive sentiment in the long term. After transgender identity disclosures on Facebook, an important means of mass disclosure, those with supportive networks experienced increased positive sentiment.

Conclusions: With foreknowledge of sentiment patterns likely to occur during gender transition, transgender people and their mental healthcare professionals can prepare with proper support in place throughout the gender transition process. Social media are a novel data source for understanding transgender people's sentiment patterns, which can help reduce mental health disparities for this marginalized population during a particularly difficult time.

Key words: transgender persons, minority health, health status disparities, mental health, social media

INTRODUCTION

Transgender people (those whose gender differs from the gender they were assigned at birth,[1] including nonbinary people) face substantial mental health disparities as compared with the general population.[2,3] In a recent survey of U.S. transgender people, 39% of respondents reported experiencing serious psychological distress in the past month, a figure approximately 8 times higher than the general population.[2] Transgender people face unique stressors and vulnerabilities including pervasive discrimination, prejudice, rejection, and violence.[2,4–8]

The research literature overwhelmingly shows that gender transition improves transgender people's well-being.[9–13] Yet, the path to improved well-being is not always direct, as minority stressors—minority stress is a type of mental distress resulting from stigma,

prejudice, and other factors specifically related to a minority identity[14,15]—can lead to more depression and anxiety during transition.[5,16] Gender transition is particularly volatile because several effects occur simultaneously: a transition into the gender and identity that matches one's internal self clearly has positive mental health effects. At the same time, as one transitions, they may face discrimination and harassment, which have negative mental health effects.[2,17]

The gender transition process is also a precarious time for transgender people's mental health because transitioning requires disclosing one's transgender identity to others.[18] Coming out as transgender can allow access to resources and support,[5] but also often involves communicating sensitive information to people who may not understand transgender identity and may be discriminatory.[19] As such, during transition many transgender people constantly balance visibility with the discriminatory conditions that may accompany disclosure.[19] Given disclosure's critical role in the gender transition process, it is an important factor in this study.

The present study's objective is to further understand patterns in the complicated interplay between gender transition and well-being using a new data source: social media. Sentiment over time during gender transition was examined using computational sentiment analysis of 41 066 posts from 240 transition blogs on the social media site Tumblr. These are blogs where transgender people document personal accounts of their experiences. This work highlights how social media data can be used to understand people's sentiment over time during gender transition. Sentiment patterns provide vital information to help mental health providers and people beginning gender transition anticipate and put support structures in place for transition's most difficult times. With this new understanding of sentiment patterns during gender transition, steps can be taken to decrease mental health disparities faced by a substantially marginalized population.

Transgender issues in medical informatics

A small body of previous research has examined transgender issues in medical and health informatics.[20–25] Most relevant to the current study, Blotner and Rajunov[26] described how transgender healthcare providers can engage with transgender communities using social media platforms to educate themselves, inform research and medical practice, and improve healthcare quality. The present study expands on this work by using social media data to illuminate gender transition sentiment patterns, which can be of great use to medical informatics researchers and transgender healthcare providers.

Disclosure and social support's relationships with well-being

Self-disclosure has been found to lead to improved mental health, physical health, and self-esteem,[27] and is necessary to receive social support.[28] Yet disclosure of stigmatized identities may also increase anxiety due to the unpredictable responses one may receive.[27] The relationship between disclosure and mental health depends greatly on the reaction one receives—supportive reactions lead to greater disclosure benefits.[29–31]

Social support is widely found to have a moderating effect on the relationship between transition status and mental health. That is, social support from one's network can mitigate the negative effects of stressors like discrimination and harassment.[10,16,32,33] How one's audience responds to a transgender identity disclosure can have a major impact on a transgender person's well-being and on ongoing social relationships.[34–36] Relationships among transgender identity

disclosure, social support, and sentiment have not yet been examined using social media data—an important undertaking because social media is increasingly pervasive in people's lives.[37]

Computational text analysis methods

Computational text analysis methods such as sentiment analysis can be powerful tools for researchers to extract meaning and themes from large bodies of text.[38,39] People who experience more psychological distress may also be more likely to post depressive content on social media;[40] thus there is an established correlation between mental health and social media content. Researchers have used computational linguistic techniques like sentiment analysis, sometimes paired with social media data, to understand social phenomena such as depression,[41–44] mental health more broadly,[45–47] and even transgender topics.[48–50] The present study uses these methods to uncover sentiment patterns over time during gender transition.

Data source: Tumblr transition blogs

Transition blogs are a genre of Tumblr blog in which people document their gender transition. Commonly, these blogs include diary-like entries discussing social, medical, and legal aspects of transition: discussion of the coming out process and resulting support or rejection, physical and mental changes, medical procedures, and name and document changes. Tumblr was chosen as the data source because it is a primary space where transgender people wrote lengthy, meaningful content about their personal experiences and emotions during transition,[51] with the added benefit of being a data source that is relatively easy for researchers to collect to gain insights into this population's sentiment. However, recent policy changes banning "adult" content have made the site substantially less welcoming for transgender people, given that transition blogs sometimes include graphic transition-related content.[52]

MATERIALS AND METHODS

Methods are summarized in Figure 1. This study was approved by the University of California, Irvine, Institutional Review Board. Parts of this work draw from a larger study.[53] Some methodological detail is omitted due to word limit; please see Haimson[53] for full methods.

Data collection

Using Tumblr's application programming interface (API)[54] and the PyTumblr API client,[55] 41 066 text or photo caption posts were collected from 240 transition blogs starting with each blog's first post. This data collection approach appears to be in line with Tumblr's API License Agreement (circa January 2017).[59] Data collection and inclusion criteria are detailed in a STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) diagram[60] in Figure 2. Data collection did not include photos, images, or visual content of any kind.

Measuring sentiment

Computational sentiment analysis was used to measure sentiment in each blog post. LIWC Linguistic Inquiry Word Count (LIWC) is a set of lexicons that enables researchers to computationally analyze people's feelings and affect via text.[56] LIWC positive emotion and negative emotion measures were assigned to each post and averaged over time to use as outcome variables in regression models. While computational sentiment measures have been found to be somewhat

Table 1. Transgender identity disclosure audiences identified in Tumblr posts

Disclosure Audience	Excerpt From Example Post (Support Code)	Post Count
Work	"As I mentioned the other day, today was the big day for me at work. Mass disclosure that I'm a transgender woman and would be transitioning at work, right in front of 500 people that I've worked with as a male for 15 years." (support unknown)	77
Stranger/acquaintance	"I only see her once a year, at con, for about 7 hours or so total. When I told her I was transitioning (had to use female ID for badge), she gave me a hug, and said she was so glad that I was happy." (supportive)	65
Friend	"I came out to my best friend tonight. I just basically mentioned it again very vaguely when we went to [restaurant] after… getting ready for the trip this weekend and I really couldn't bear to just continue dancing around it and being so vague when I knew how accepting he would be. So I just told him. Really awesome to the point that I can't describe it! He basically told me he has my back continually and if it comes to my worst case scenario then he will continue to support me and help me out. That's fucking huge and I can't really state how happily excited I am." (supportive)	57
Extended family	"I came out to my extended family yesterday, and I've heard back from all but one (the conservative one, hmmm). And they're all accepting and supportive!!! This makes me so happy!! Yay!!!!" (supportive)	56
Mom	"I came out to my mother today. She cried. And then she told me she didn't understand why. Then she cried. Then she told me it was fine as long as I was happy. And then, yes, she cried again. Overall I think it went well." (partially supportive)	35
Sibling	"I came out to my brother today (he's fourteen). He was the last important person I had to tell. And he was very cool with it. I asked him if I felt kind of like a brother to him and he said yes." (supportive)	29
Dad	"Emailed my short coming out note to dad. Hope it goes well." (support unknown)	28
Facebook	"I just came out to my friends on Facebook shit I'm shaking and kinda terrified but also feeling good at the same time." (support unknown)	26
School	"I came out to my faculty and supervisors at school and this was a bit trickier to do. I decided to wait until right before the semester started as I felt this was a good time to begin something new. Right before the new year :) Sometimes it is hard to come out to people professionally because being trans is a very personal matter. You also never know how anyone in the education system may react… In the end the support was even more overwhelming." (supportive)	18
Unknown	"Well this is the most awkward and awful coming out conversation I've ever had. :|" (not supportive)	15
Everyone	"So my birthday was yesterday… and I came out to everyone!!! It felt wonderful so much lifted off of my shoulders. I know it's a long journey but I now know I have a lot more support than I thought and I'm so ready to truly be me and not have to hide!" (supportive)	13
Health professional	"So the tech that does the laser hair removal is extremely nice, and very familiar with the transgender MtF community… She's the third person I've come out to so far. She congratulated me and gave me nothing but smiles." (supportive)	11
Past acquaintance	"That awkward moment where you take a delivery to someone who you knew back in high school. You know, before transition." (support unknown)	10
Romantic interest	"Anyways I ended up meeting a boy there that was pretty cute who was giving me lots of looks and we ended up spending all night together. Until he had to give a friend a ride home. Well he gave me his number, and knew if I was going to text him I would have to come out to him right away. Of course you could tell he was really interested until I told him. Then he immediately told me he wasn't interested. (not supportive)	8
Child	excerpt not included	4
Church	excerpt not included	3
Partner	excerpt not included	3
Ex-partner	excerpt not included	2
Instagram	excerpt not included	2
Twitter	excerpt not included	2
Total		362[a]

Blog post quotes were not traceable via Google search as of March 2019, and so were left as is; otherwise, they would have been paraphrased to reduce traceability to maintain bloggers' privacy.

[a]Total is not a sum of the rows because many disclosure posts had multiple audiences.

of a proxy for emotional well-being,[39,61] these measures have accuracy limitations[39,61] and are not a clinical measure of mental health.

A machine learning classifier to detect transgender identity disclosures

A machine learning classifier was built to detect a particular type of disclosure: Tumblr posts describing transgender identity disclosures in other contexts (see Table 1 for example posts). A post counted as a transgender identity disclosure if it described a disclosure that seemed to have occurred within 2 weeks prior to the post describing it.

The first step involved building a training set of positive and negative examples of transgender identity disclosure posts in the dataset. An iterative approach was used to build a sufficient training set, which included several rounds of manual coding and machine learning. To establish interrater reliability, 2 coders (OLH and NA) first coded 50 posts as either recent transgender identity disclosures or not, and reached acceptable interrater agreement at a kappa of 0.72. OLH then coded the remaining training data. The Python SciKitLearn library[62] was used to build the machine learning classifier. The classifier's features are detailed in Figure 1. Nine machine learning algorithms were experimented: AdaBoost, decision tree, k-
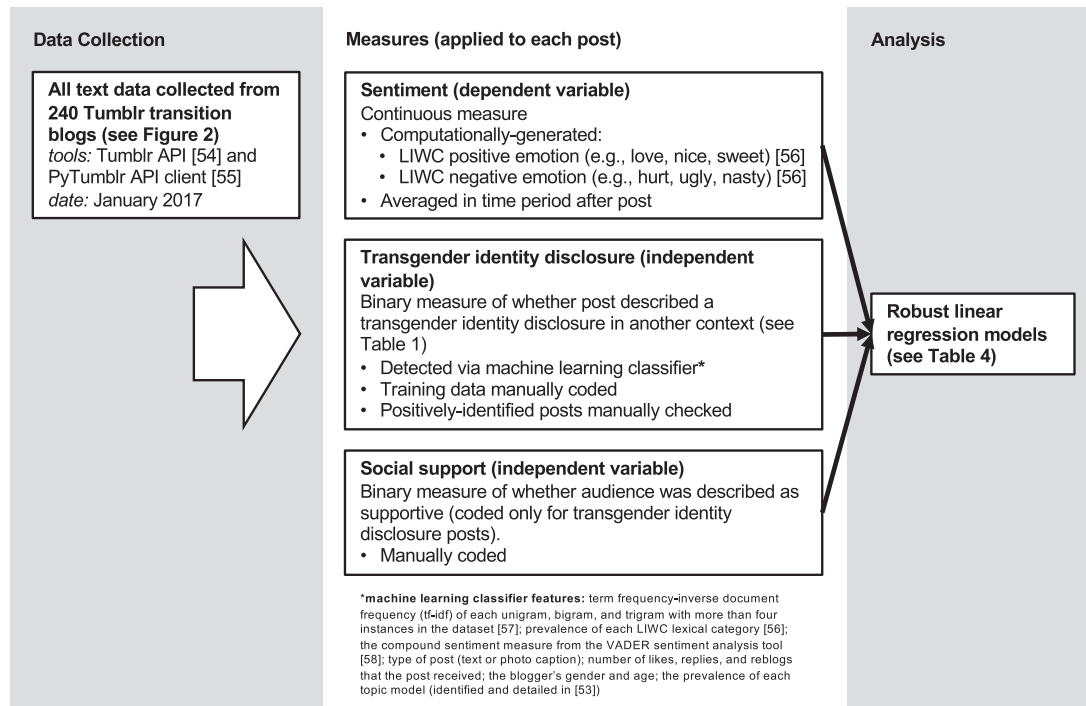
Figure 1. Methods summary.

nearest neighbors, logistic regression, naïve Bayes (Bernoulli, Multinomial, and Gaussian), random forest, and support vector classification. AdaBoost was most accurate, with an accuracy of 0.80 and area under the curve of 0.62 when applying 10-fold cross-validation. When applied to the 20% of data held out as a test set, the classifier's accuracy was 0.79 and the area under the curve was 0.71.

Next the classifier was applied to the full dataset. The model classified 798 posts as positive, which OLH then manually coded to ensure that the computational coding did not include false positives. Manual coding identified a total of 362 posts describing recent transgender identity disclosures. The high number of false positives indicates that the model had poor specificity, a limitation that was addressed by manually coding all positively classified posts. Unfortunately, it is not possible to identify false negatives. For each transgender identity disclosure post, the disclosure audience(s) was manually identified by reading the post. This resulted in a set of 20 disclosure audience types (Table 1).

Measuring social support

Each post that described a transgender identity disclosure was manually coded for whether the poster described their audience as being supportive in response to the disclosure (yes, no, partially, or unknown). This was later simplified to a binary variable (supportive response or not) after observing few posts in the partially and unknown categories.

Understanding relationships among sentiment, transgender identity disclosures, and social support

As a result of the previous 3 steps, each post in the dataset had the following information:

1. variables measuring the post's positive and negative sentiment (**dependent variables**)

2. whether or not the post described a recent transgender identity disclosure (0 or 1) (**independent variable**)

   Additionally, transgender identity disclosure posts had a measure of the following:

3. whether the disclosure received a supportive response (0 or 1) (**independent variable**)

Regression models were built to understand the relationships between these variables. Using posts as the unit of analysis, all models include average sentiment in the time period after the post (1-30 days, 1-90 days, or 1-180 days) as the dependent variable. Independent variables included whether or not the post described a transgender identity disclosure, and whether or not the disclosure received a positive response. The models also included all available control variables, including blogger demographics and characteristics of posts. An important control variable was previous sentiment, because a person's sentiment one month is highly predictive of their sentiment the next month; yet even after controlling for this, the variables of interest still show significant effects. Because the data did not meet the assumptions required for linear regression, robust linear regression was used instead.[63] Data cleaning procedures are detailed in Haimson.[53]

Mapping the gender transition process into stages

In previous work, qualitative analysis (based on content analysis of Tumblr blog data and interviews with bloggers) was used to map the gender transition process into stages based on van Gennep's liminality framework.[51,53,64] In section R.1, all transgender identity disclosures are grouped together, regardless of audience. However, because it is more informative to separate out different types of disclosures, this is done according to the primary disclosures that related to the liminality stages:[51,53,64] family (section R.2) and Facebook (section R.3). Facebook disclosures were generally mass disclosures to a broad set of people in one's Facebook network.[51]
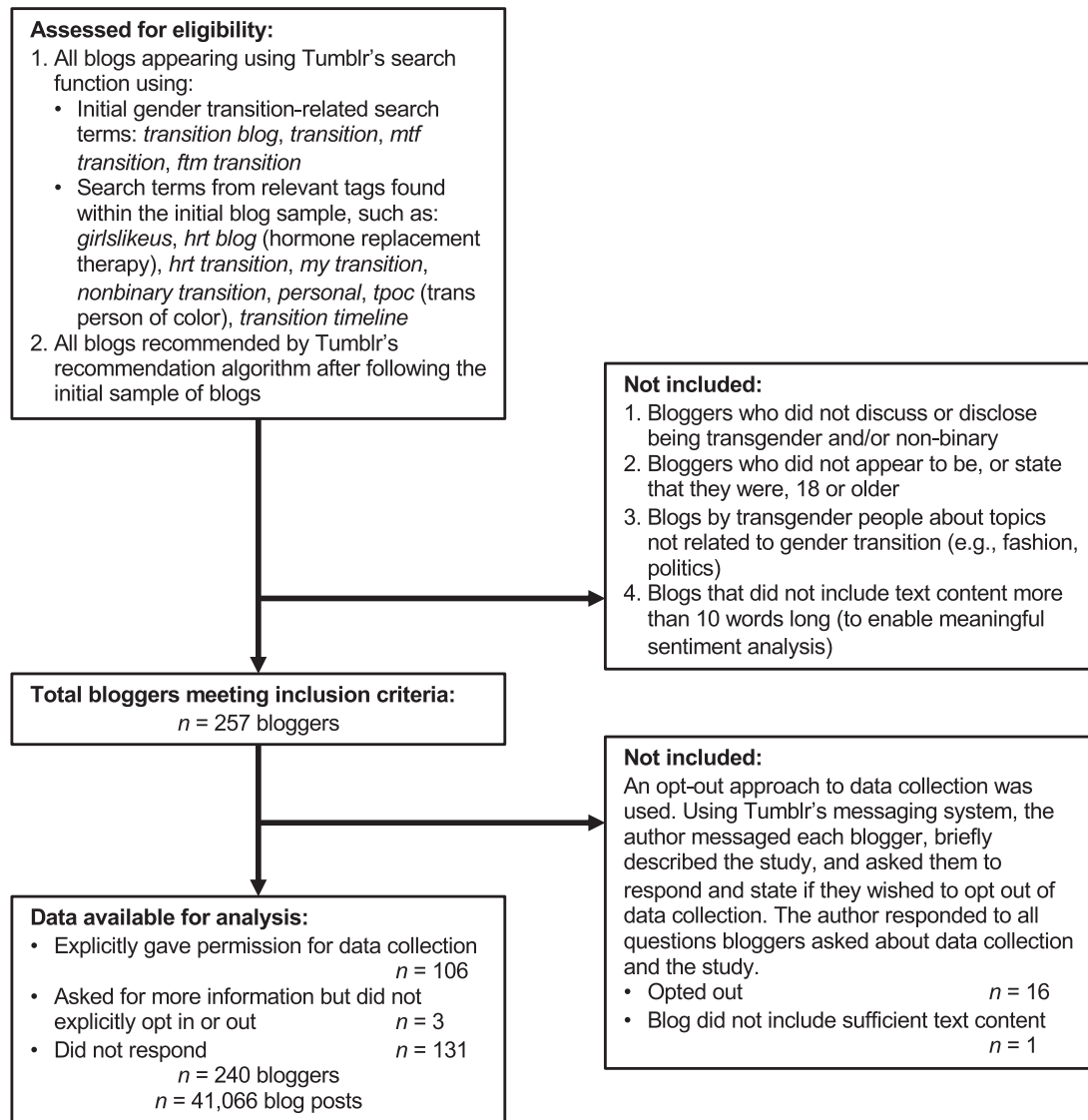
Assessed for eligibility:
1. All blogs appearing using Tumblr's search function using:
   - Initial gender transition-related search terms: *transition blog, transition, mtf transition, ftm transition*
   - Search terms from relevant tags found within the initial blog sample, such as: *girlslikeus, hrt blog* (hormone replacement therapy), *hrt transition, my transition, nonbinary transition, personal, tpoc* (trans person of color), *transition timeline*
2. All blogs recommended by Tumblr's recommendation algorithm after following the initial sample of blogs

Not included:
1. Bloggers who did not discuss or disclose being transgender and/or non-binary
2. Bloggers who did not appear to be, or state that they were, 18 or older
3. Blogs by transgender people about topics not related to gender transition (e.g., fashion, politics)
4. Blogs that did not include text content more than 10 words long (to enable meaningful sentiment analysis)

Total bloggers meeting inclusion criteria:
n = 257 bloggers

Not included:
An opt-out approach to data collection was used. Using Tumblr's messaging system, the author messaged each blogger, briefly described the study, and asked them to respond and state if they wished to opt out of data collection. The author responded to all questions bloggers asked about data collection and the study.
- Opted out   *n* = 16
- Blog did not include sufficient text content   *n* = 1

Data available for analysis:
- Explicitly gave permission for data collection   *n* = 106
- Asked for more information but did not explicitly opt in or out   *n* = 3
- Did not respond   *n* = 131
   n = 240 bloggers
   n = 41,066 blog posts

Figure 2. Data collection and inclusion criteria flow diagram for Tumblr transition blogs.

Table 2. Sentiment changes over time on average after transgender identity disclosures

	Short Term	Long Term
All disclosures (R.1) (regardless of audience)	decreased negative sentiment if supportive response	increased positive sentiment
Family disclosures (R.2)	increased negative sentiment	increased positive sentiment
Facebook disclosures (R.3)	increased positive sentiment if supportive response	increased positive sentiment if supportive response

RESULTS

Results are summarized in Table 2. Table 3 provides clarity about the organization of models in Table 4. Results are detailed in Table 4 and sections R.1-R.3.

Data description

A total of 41 066 Tumblr posts by 240 bloggers were analyzed. Posts had an average word count of 71.38 (median = 33, SD = 124.70). On average, each blogger posted 367 total text posts that met the data collection criteria (median = 76, SD = 814.46), had been blogging for almost 2 years (mean = 646 ± 515.19 days, median = 530 days), and posted roughly 3 times/week (mean = 0.43 ± 0.67 posts/day). Demographic data were found in blog descriptions. Most (95%) prominently stated or implied their gender and many (42%) stated their age. When placing each blogger into the most prominent gender category that they displayed on their blog (with the caveat that some identified as more than 1 gender), bloggers in the dataset were 47% trans men, 46% trans women, and 7% nonbinary (including genders such as genderqueer, genderfluid, and agender). Like Tumblr more broadly,[65] the sample skewed young, with 63% in the 18-24 years of age range, 30% were 25-34 years of age, 7% were 35-44 years of age, and <1% were 45 years of age or older. Most bloggers (93%) did not specify race or ethnicity. Posts were from 2009 to 2017 (collected in

Table 3. Description of regression models

		Time Period After Post Averaged for Outcome Variable		
		Days 1-30	Days 1-90	Days 1-180
Sentiment measured in outcome variable	Negative sentiment	Models 1, 7, 13	Model 2, 8, 14	Model 3, 9, 15
	Positive sentiment	Model 4, 10, 16	Model 5, 11, 17	Model 6, 12, 18

early January 2017), with 83.7% being from 2014 to 2016. Bloggers described on average 1.51 transgender identity disclosures each (median = 0, SD = 2.98, maximum = 18).

R.1 Transgender identity disclosures are followed by decreased negative sentiment and increased positive sentiment

Support moderates the relationship between disclosure and negative sentiment in the short term. Model 1 (Table 4) shows that posts describing transgender identity disclosures with supportive responses were followed by fewer negative emotion words in the month following.

Disclosure is associated with increased positive sentiment in the long term, whether or not the audience was supportive. Models 5 and 6 (Table 4) indicate that posts describing transgender identity disclosures were followed by more positive emotion words in the 3 and 6 months following, but support was not a moderating variable. That is, people saw positive sentiment increases in the 3 and 6 months post disclosure whether or not their disclosure audience responded supportively. Outcome measures are the percentage of total words in a post that were part of the LIWC positive emotion dictionary. Thus, Model 5's coefficient of 0.25 means that posts in the 3 months after transgender identity disclosures had on average 0.25% more positive emotion words. This level of detail will not be provided for each result, but is provided here to help readers interpret Table 4.

R.2 Family disclosures are followed by increased negative sentiment in the short term, but increased positive sentiment in the long term

Here, the independent variable of interest is a binary indicator of whether a post described a recent disclosure to a family member. These involved either specific family members (eg, parent, sibling, child, grandparent), or a mention of family more broadly (eg, "I came out to my family today!").

Family disclosures are associated with increased negative sentiment in the short term. Posts describing transgender identity disclosures to family members were followed by more negative emotion words in the month following, according to Model 7 (Table 4). Importantly, support is not a significant moderating variable here; even those who received supportive responses from family members experienced increased negative short term sentiment.

Family disclosures are associated with increased positive sentiment in the long term. Model 12 (Table 4) shows that posts describing transgender identity disclosures to family members were followed by more positive emotion words in the 6 months following. Again, support was not a moderating variable. Positive sentiment increased whether or not people received positive responses after disclosing to family members.

R.3 Facebook disclosures with supportive responses are followed by increased positive sentiment

Because transgender identity disclosures on Facebook are pivotal transition experiences,[51] it is important to understand how people's sentiment changed afterwards. The independent variable of interest is a binary indicator of whether a Tumblr post described a Facebook disclosure.

Support moderates the relationship between Facebook disclosures and positive sentiment in the short term and long term. Models 16-18 (Table 4) indicate that posts describing transgender identity disclosures on Facebook with supportive responses were followed by more positive emotion words in both the short term and the long term. None of the models show a direct significant relationship between Facebook disclosures and sentiment. It makes sense that a person's sentiment after disclosing on Facebook is highly dependent on the response that they receive from their networks. Support from one's network is an important moderating variable impacting people's positive sentiment after disclosing on Facebook.

DISCUSSION

Sentiment patterns throughout gender transition

This work contributes an understanding of the patterns in sentiment changes throughout gender transition. Figure 3 displays a conceptual visualization of these patterns over time (with disclosure processes as mapped onto van Gennep's 3 liminality stages[64] from previous work).[53] During the separation stage, which involves disclosures to family members, sentiment decreases on average. Next, during the transition stage, which involves transgender identity disclosures on Facebook, positive sentiment increases. Finally, in the incorporation stage, people's positive sentiment increases to a level on average higher than their pretransition positive sentiment. The initial decreased sentiment shown in Figure 3 is likely a result of the combined impact of family disclosures along with minority stressors like discrimination, harassment, and disapproval from others, and personal discomfort in the early stages of transition. The long-term increased positive sentiment may correspond to people's bodies and social identities aligning with their internal gender, and others in their lives becoming more supportive over time. These 2 effects happen in tandem, but results indicate that (on average) in the short term the difficulties and negative aspects of transition are more prominent, while in the long term, transition's positive benefits prevail.

The short-term increase in negative sentiment after transgender identity disclosures to family members is a surprising result, given that previous literature has found that disclosures are generally followed by positive emotions.[27,66,67] Yet similar results to the present study's have been found in the context of schizophrenia disclosures on Twitter[68] and emotional writing:[69] in many cases, sensitive disclosures are followed by increased negative affect in the short term, and the positive benefits of disclosure take time to occur. Interview data as part of the larger study[53] indicate that increases in negative sentiment after family disclosures are sometimes in response to family disclosures. However, negative sentiment increases likely also relate to the broader difficulties and minority stressors[5] that people face during the stage of gender transition that corresponds tempo-

Table 4. Robust linear regression models showing average negative and positive sentiment in time period following posts

Variable (binary indicators)	Average LIWC Negative Emotion in Days 1-30 After Post		Average LIWC Negative Emotion in Days 1-90 After Post		Average LIWC Negative Emotion in Days 1-180 After Post		Average LIWC Positive Emotion in Days 1-30 After Post		Average LIWC Positive Emotion in Days 1-90 After Post		Average LIWC Positive Emotion in Days 1-180 After Post	
All disclosures (regardless of audience) (n = 28 968 posts), Section R.1	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Disclosure	0.12	(0.08)	0.01	(0.06)	-0.03	(0.05)	0.03	(0.11)	$\mathbf{0.25^c}$	(0.08)	$\mathbf{0.24^d}$	(0.07)
Supportive response from disclosure audience	-0.22^b	(0.11)	-0.11	(0.08)	-0.06	(0.07)	-0.12	(0.15)	-0.11	(0.10)	-0.04	(0.09)
Intercept	1.77^d	(0.03)	1.21^d	(0.02)	1.05^d	(0.02)	1.84^d	(0.04)	1.27^d	(0.03)	1.06^d	(0.03)
R^2	0.34		0.54		0.62		0.27		0.45		0.51	
Family disclosures (n = 33 173 posts), Section R.2	Model 7		Model 8		Model 9		Model 10		Model 11		Model 12	
Disclosure to family member	$\mathbf{0.36^c}$	(0.12)	0.11	(0.09)	0.03	(0.08)	-0.22	(0.17)	0.13	(0.12)	$\mathbf{0.21^b}$	(0.10)
Supportive response from family	-0.16	(0.18)	-0.09	(0.13)	-0.05	(0.12)	-0.13	(0.25)	-0.15	(0.18)	-0.14	(0.15)
Disclosure to other audience	0.08	(0.10)	0.02	(0.07)	0.02	(0.07)	0.16	(0.14)	$\mathbf{0.27^c}$	(0.10)	$\mathbf{0.18^b}$	(0.09)
Supportive response from others	-0.31^b	(0.15)	-0.24^b	(0.11)	-0.14	(0.10)	-0.29	(0.21)	-0.18	(0.14)	-0.01	(0.13)
Intercept	1.72^d	(0.03)	1.17^d	(0.02)	1.01^d	(0.02)	1.89^d	(0.04)	1.28^d	(0.03)	1.06^d	(0.03)
R^2	0.34		0.54		0.61		0.27		0.45		0.51	
Facebook disclosures (n = 35 346 posts), Section R.3	Model 13		Model 14		Model 15		Model 16		Model 17		Model 18	
Facebook disclosure	-0.00	(0.27)	0.00	(0.19)	0.03	(0.17)	0.00	(0.37)	0.20	(0.26)	0.26	(0.23)
Supportive response on Facebook	-0.47	(0.41)	0.05	(0.30)	-0.02	(0.26)	$\mathbf{1.52^c}$	(0.57)	$\mathbf{0.91^b}$	(0.40)	$\mathbf{0.71^b}$	(0.36)
Disclosure to other audience	$\mathbf{0.15^b}$	(0.07)	0.07	(0.05)	0.00	(0.05)	-0.03	(0.10)	$\mathbf{0.16^b}$	(0.07)	$\mathbf{0.16^b}$	(0.06)
Supportive response from others	-0.17	(0.11)	-0.12	(0.08)	-0.05	(0.07)	-0.17	(0.15)	-0.09	(0.11)	-0.01	(0.09)
Intercept	1.81^d	(0.03)	1.21^d	(0.02)	1.06^d	(0.02)	1.87^d	(0.04)	1.27^d	(0.03)	1.04^d	(0.03)
R^2	0.34		0.54		0.61		0.26		0.44		0.50	

Values are coefficient (SE). Control variables (included in each model; details omitted for space): involuntary disclosure (binary indicator), gender, age, number of likes, number of replies, number of reblogs, word count, year, average negative emotion in time period before post, average positive emotion in time period before post. Please see Supplementary Material for full regression tables.
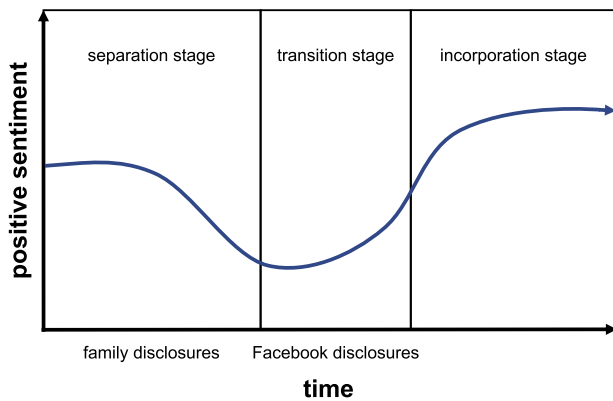
LIWC: Linguistic Inquiry Word Count.

[a] $p < .10$;
[b] $p < .05$;
[c] $p < .01$;
[d] $p < .001$.

Bolded values indicate significant results discussed in sections R.1–R.3.

Figure 3. Conceptual visualization of gender transition sentiment patterns over time on average.

rally to family disclosures,[53] or fear and anxiety about upcoming disclosures to other audiences. Considering this result in the context of overall gender transition sentiment patterns provides clarity.

Implications for healthcare providers and transgender individuals

Understanding patterns in sentiment over time during gender transition has important implications for helping a marginalized population—transgender people—find the support they need at the times they most need it, which can decrease the substantial mental health disparities this population faces. This study's results provide important information for (1) mental health professionals who wish to support their patients and (2) those beginning gender transitions themselves to anticipate changes in emotional state and put support structures in place to help navigate the difficult stages of transition.

Mental healthcare providers working with transgender populations, and transgender people in transition, are both important audience for patterns identified by analyzing sentiment over time via social media data. Understanding patterns in advance can help mental healthcare providers better anticipate and provide the support their patients need at the times most needed. Health professionals are often some of the very first people to whom trans people disclose their trans identity.[53] Blotner and Rajunov[26] argued that medical professionals serving transgender patients can use social media to understand trans lived experiences, and the present argument takes this a step further. Equipped with average transition sentiment trajectories harnessed from social media data, therapists could help their patients know what to expect, and ensure that they have sufficient support structures in place, as they embark on gender transition and transgender identity disclosures to the people in their lives. In addition to the aggregate work presented here, future work could examine how people's individual sentiment patterns may be identified from social media to provide important insights for their mental healthcare providers. Researchers can be the important links between social media data, populations in need of support, and healthcare professionals capable of providing support when provided this extra information. That said, privacy protections must be ensured in every step of the data access chain from patients' social media data to healthcare professional's data access.

Implications for researchers

This research highlights how social media is an important data source for quickly and effectively understanding sentiment over time

throughout gender transition, a method which researchers can also apply to other populations facing health disparities, such as racial or ethnic minorities and people with specific medical conditions. Other studies[10,16] have primarily measured emotional well-being throughout gender transition using research methods like surveys and interviews. Though computational methods have accuracy limitations,[39,61] self-reported data sources like surveys and interviews can be faulty given people's difficulty recalling emotion in the pasts[70] or even interpreting their current emotional state. Additionally, traditional research procedures, particularly in longitudinal studies, are often costly and time consuming, both for researchers and for research populations. Marginalized populations, particularly those facing health disparities, may not have the time, energy, or willingness to fill out surveys or participate in interviews or longitudinal studies. If done ethically, harnessing social media data—a data source that populations are already using to chronicle their experiences—is a powerful method for uncovering insights about sentiment over time.

Limitations

Beyond the accuracy limitations of computational sentiment detection methods and machine learning disclosure detection methods noted above, several other limitations arise. First, this study's sample is not representative of all transgender people. Next, there may have been additional control variables that confound the relationship between sentiment and disclosure. Finally, while it is important to understand average sentiment patterns over time during gender transition, it is also important to consider that averages are only that: averages. The data in this study represent people's lives, meaning that patterns are messy and involve high variance from person to person. Average sentiment over time patterns are complicated by the intersecting identity facets and other life transitions that people experience at the same time as their gender transitions. Those who have other stigmatized identity facets, and who experience other distressing life changes along with gender transition, have very different sentiment trajectories over time. This is an important area to examine in future work.

CONCLUSION

This research shows how people's sentiment changed over time on average during gender transition using social media data. Findings suggest that while overall people's positive sentiment increased throughout transition and after disclosing their transgender identity to the people in their lives, sentiment was highly dependent on whether people received support from their disclosure audiences. In the short term after family disclosures (during the difficult early stages of transition), people's sentiment suffered whether or not their families were supportive. Gender transition is difficult, yet understanding sentiment patterns in advance can help to mitigate some of these difficulties. This work provides empirical knowledge so that people in transition, along with their mental health professionals and support structures, can have foreknowledge of the patterns likely to occur, and thus can be prepared to improve the experience. In this way, this research contributes to lessening the mental health disparities widely faced by transgender people.

FUNDING

AUTHOR CONTRIBUTORS

OLH conducted data collection, data analysis, and writing.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Stryker S. *Transgender History*. Berkeley, CA: Seal Press; 2009.
2. James SE, Herman JL, Rankin S, *et al*. *The Report of the 2015 U.S. Transgender Survey*. Washington, DC: National Center for Transgender Equality; 2016. http://www.transequality.org/sites/default/files/docs/USTS-Full-Report-FINAL.PDF. Accessed December 8, 2016.
3. Reisner SL, White JM, Bradford JB, *et al*. Transgender health disparities: comparing full cohort and nested matched-pair study designs in a community health center. *LGBT Health* 2014; 1 (3): 177–84.
4. Clements-Nolle K, Marx R, Katz M. Attempted suicide among transgender persons. *J Homosex* 2006; 51 (3): 53–69.
5. Hendricks ML, Testa RJ. A conceptual framework for clinical work with transgender and gender nonconforming clients: an adaptation of the Minority Stress Model. *Prof Psychol Res Pract* 2012; 43 (5): 460–7.
6. Lombardi EL, Wilchins RA, Priesing D, *et al*. Gender violence: transgender experiences with violence and discrimination. *J Homosex* 2002; 42 (1): 89–101.
7. Rood BA, Reisner SL, Surace FI, *et al*. Expecting rejection: understanding the minority stress experiences of transgender and gender-nonconforming individuals. *Transgender Health* 2016; 1 (1): 151–64.
8. Rood BA, Puckett JA, Pantalone DW, *et al*. Predictors of suicidal ideation in a statewide sample of transgender individuals. *LGBT Health* 2015; 2 (3): 270–5.
9. What We Know: The Public Policy Research Portal, Center for the Study of Inequality at Cornell University. What does the scholarly research say about the effect of gender transition on transgender well-being? https://whatweknow.inequality.cornell.edu/topics/lgbt-equality/what-does-the-scholarly-research-say-about-the-well-being-of-transgender-people/. Accessed January 4, 2019.
10. Budge SL, Adelson JL, Howard K. Anxiety and depression in transgender individuals: The roles of transition status, loss, social support, and coping. *J Consult Clin Psychol* 2013; 81 (3): 545–57.
11. Durwood L, McLaughlin KA, Olson KR. Mental health and self-worth in socially transitioned transgender youth. *J Am Acad Child Adolesc Psychiatry* 2017; 56 (2): 116–23.e2.
12. Kozee HB, Tylka TL, Bauerband LA. Measuring transgender individuals' comfort with gender identity and appearance development and validation of the transgender congruence scale. *Psychol Women Q* 2012; 36 (2): 179–96.
13. Strain JD, Shuff IM. Psychological well-being and level of outness in a population of male-to-female transsexual women attending a national transgender conference. *Int J Transgenderism* 2010; 12 (4): 230–40.
14. Brooks VR. *Minority Stress and Lesbian Women*. Lexington, MA: Lexington Books; 1981.
15. Meyer IH. Minority stress and mental health in gay men. *J Health Soc Behav* 1995; 36 (1): 38–56.
16. Bockting WO, Miner MH, Swinburne Romine RE, *et al*. Stigma, mental health, and resilience in an online sample of the US transgender population. *Am J Public Health* 2013; 103 (5): 943–51.
17. Grant JM, Mottet LA, Tanis JD, Harrison J, Herman JL, Keisling M. *Injustice at Every Turn: A Report of the National Transgender Discrimination Survey*. Washington, DC: National Center for Transgender Equality and National Gay and Lesbian Task Force; 2011.
18. Haimson OL, Brubaker JR, Dombrowski L, *et al*. Disclosure, stress, and support during gender transition on Facebook. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. New York, NY: ACM; 2015: 1176–90.
19. Levitt HM, Ippolito MR. Being transgender navigating minority stressors and developing authentic self-presentation. *Psychol Women Q* 2014; 38 (1): 46–64.
20. Blosnich JR, Cashy J, Gordon AJ, *et al*. Using clinician text notes in electronic medical record data to validate transgender-related diagnosis codes. *J Am Med Inform Assoc* 2018; 25 (7): 905–8.
21. Deutsch MB, Green J, Keatley J, *et al*. Electronic medical records and the transgender patient: recommendations from the World Professional Association for Transgender Health EMR Working Group. *J Am Med Inform Assoc* 2013; 20 (4): 700–3.
22. Grasso C, McDowell MJ, Goldhammer H, *et al*. Planning and implementing sexual orientation and gender identity data collection in electronic health records. *J Am Med Inform Assoc* 2019; 26 (1): 66–70.
23. Duncan J, Narus SP, Clyde S, *et al*. Birth of identity: understanding changes to birth certificates and their value for identity resolution. *J Am Med Inform Assoc* 2015; 22: e120–9.
24. Deutsch MB. Evaluation of patient-oriented, internet-based information on gender-affirming hormone treatments. *LGBT Health* 2016; 3 (3): 200–7.
25. Marshall Z, Brunger F, Welch V, *et al*. Open availability of patient medical photographs in google images search results: cross-sectional study of transgender research. *J Med Internet Res* 2018; 20 (2): e70.
26. Blotner C, Rajunov M. Engaging transgender patients: using social media to inform medical practice and research in transgender health. *Transgender Health* 2018; 3 (1): 225–8.
27. Rosario M, Hunter J, Maguen S, *et al*. The coming-out process and its adaptational and health-related associations among gay, lesbian, and bisexual youths: stipulation and exploration of a model. *Am J Community Psychol* 2001; 29 (1): 133–60.
28. Chaudoir SR, Fisher JD. The disclosure processes model: understanding disclosure decision making and postdisclosure outcomes among people living with a concealable stigmatized identity. *Psychol Bull* 2010; 136 (2): 236–56.
29. Lepore SJ, Ragan JD, Jones S. Talking facilitates cognitive–emotional processes of adaptation to an acute stressor. *J Pers Soc Psychol* 2000; 78 (3): 499–508.
30. Puckett JA, Woodward EN, Mereish EH, *et al*. Parental rejection following sexual orientation disclosure: impact on internalized homophobia, social support, and mental health. *LGBT Health* 2015; 2 (3): 265–9.
31. Rodriguez RR, Kelly AE. Health effects of disclosing secrets to imagined accepting versus nonaccepting confidants. *J Soc Clin Psychol* 2006; 25 (9): 1023–47.
32. Rotondi NK, Bauer GR, Scanlon K, *et al*. Prevalence of and risk and protective factors for depression in female-to-male transgender Ontarians: Trans PULSE Project. *Can J Commun Ment Health* 2011; 30 (2): 135–55.

33. Rotondi NK, Bauer GR, Travers R, *et al*. Depression in male-to-female transgender Ontarians: results from the Trans PULSE Project. *Can J Commun Ment Health* 2011; 30 (2): 113–33.

34. Galupo MP, Krum TE, Hagen DB, *et al*. Disclosure of transgender identity and status in the context of friendship. *J LGBT Issues Couns* 2014; 8 (1): 25–42.

35. Mullen G, Moane G. A qualitative exploration of transgender identity affirmation at the personal, interpersonal, and sociocultural levels. *Int J Transgenderism* 2013; 14 (3): 140–54.

36. Nuttbrock L, Bockting W, Rosenblum A, *et al*. Gender identity conflict/affirmation and major depression across the life course of transgender women. *Int J Transgenderism* 2012; 13 (3): 91–103.

37. Perrin A. Social media usage: 2005-2015. *Pew Research Center Internet and Technology*. 2015. http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/. Accessed July 31, 2016.

38. Blei DM, Lafferty JD. Dynamic topic models. In: *Proceedings of the 23rd International Conference on Machine Learning*. New York, NY: ACM; 2006: 113–20.

39. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 2010; 29 (1): 24–54.

40. Bazarova NN, Choi YH, Whitlock J, *et al*. Psychological distress and emotional expression on Facebook. *Cyberpsychology Behav Soc Netw* 2017; 20 (3): 157–63.

41. De Choudhury M, Gamon M, Counts S, *et al*. Predicting depression via social media. In: *ICWSM AAAI Conference on Weblogs and Social Media*; 2013.

42. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. *Annu Rev Psychol* 2003; 54 (1): 547–77.

43. De Choudhury M, Counts S, Horvitz EJ, *et al*. Characterizing and predicting postpartum depression from shared Facebook data. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. New York, NY: ACM; 2014: 626–38.

44. De Choudhury M, Counts S, Horvitz E. Social media as a measurement tool of depression in populations. In: *Proceedings of the 5th Annual ACM Web Science Conference*. New York, NY: ACM; 2013: 47–56.

45. De Choudhury M, De S. Mental health discourse on reddit: self-disclosure, social support, and anonymity. In: *ICWSM AAAI Conference on Weblogs and Social Media*; 2014.

46. De Choudhury M, Kiciman E, Dredze M, *et al*. Discovering Shifts to Suicidal ideation from mental health content in social media. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York, NY: ACM; 2016: 2098–110.

47. De Choudhury M, Kıcıman E. The language of social support in social media and its effect on suicidal ideation risk. In: *ICWSM AAAI Conference on Weblogs and Social Media*; 2017.

48. Haimson OL, Hayes GR. Changes in social media affect, disclosure, and sociality for a sample of transgender Americans in 2016's political climate. In: *ICWSM AAAI Conference on Weblogs and Social Media*; 2017.

49. Pennebaker JW, Groom CJ, Loew D, *et al*. Testosterone as a social inhibitor: two case studies of the effect of testosterone treatment on language. *J Abnorm Psychol* 2004; 113 (1): 172–5.

50. Karami A, Webb F, Kitzie VL. Characterizing transgender health issues in Twitter. In: *Proceedings of the Association for Information Science and Technology*; 2018.

51. Haimson OL. Social media as social transition machinery. *Proc ACM Hum-Comput Interact* 2018; 2 (CSCW): 63:1–21.

52. Broker EJ. Why Tumblr's "Adult Content" ban threatens to close a crucial space of exploration for trans women. *Slate*. 2018. https://slate.com/human-interest/2018/12/tumblr-adult-content-ban-trans-women.html. Accessed March 5, 2019.

53. Haimson OL. *The Social Complexities of Transgender Identity Disclosure on Social Media*. Irvine, CA: University of California; 2018.

54. Tumblr. Tumblr API. https://www.tumblr.com/docs/en/api/v2. Accessed December 15, 2016.

55. PyTumblr. PyTumblr. GitHub. https://github.com/tumblr/pytumblr. Accessed December 15, 2016.

56. Pennebaker JW, Boyd RL, Jordan K, *et al. The Development and Psychometric Properties of LIWC 2015*. Austin, TX: University of Texas at Austin; 2015. https://utexas-ir.tdl.org/handle/2152/31333. Accessed December 15, 2016.

57. Manning CD, Raghavan P, Schütze H. *Introduction to Information Retrieval*. New York, NY: Cambridge University Press; 2008.

58. Hutto CJ, Vader GE. A parsimonious rule-based model for sentiment analysis of social media text. In: *ICWSM AAAI Conference on Weblogs and Social Media*; 2014.

59. Tumblr. Application Developer and API License Agreement; 2014. https://www.tumblr.com/docs/en/api_agreement. Accessed July 31, 2016.

60. Vandenbroucke JP, von EE, Altman DG, *et al*. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007; 4 (10): e297.

61. Bantum EO, Owen JE. Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. *Psychol Assess* 2009; 21 (1): 79–88.

62. Pedregosa F, Varoquaux G, Gramfort A, *et al*. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011; 12: 2825–30.

63. UCLA Institute for Digital Research and Education. Robust Regression. R Data Analysis Examples. IDRE Stats. https://stats.idre.ucla.edu/r/dae/robust-regression/. Accessed August 8, 2017.

64. van Gennep A. *The Rites of Passage*. Chicago, IL: University of Chicago Press; 1909.

65. Smith C. Tumblr offers advertisers a major advantage: young users, who spend tons of time on the site. *Business Insider*. 2013. http://www.businessinsider.com/tumblr-and-social-media-demographics-2013-12. Accessed July 31, 2016.

66. Pennebaker JW. *Emotion, Disclosure & Health*. Washington, DC: American Psychological Association; 1995.

67. Pennebaker JW, Kiecolt-Glaser JK, Glaser R. Disclosure of traumas and immune function: health implications for psychotherapy. *J Consult Clin Psychol* 1988; 56 (2): 239–45.

68. Ernala SK, Rizvi AF, Birnbaum ML, *et al*. Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. *Proc Acm Hum-Comput Interact* 2017; 1 (CSCW): 43.

69. Baikie KA, Wilhelm K. Emotional and physical health benefits of expressive writing. *Adv Psychiatr Treat* 2005; 11 (5): 338–46.

70. Thomas DL, Diener E. Memory accuracy in the recall of emotions. *J Pers Soc Psychol* 1990; 59 (2): 291.