Detecting Expressive Writing in Online Health Communities by Modeling Aggregated Empirical Data

HAIWEI MA, University of Minnesota, USA
SUNNY PARAWALA, University of Minnesota, USA
SVETLANA YAROSH, University of Minnesota, USA

Although expressive writing has proved to be beneficial on physical, mental, and social health of individuals, it has been restrained to lab-based experimental studies. In the real world, individuals may naturally engage with expressive writing when dealing with difficult times, especially when facing a tough health journey. Health blogging may serve as an easy-to-access method for self-therapy, if spontaneous expressive writing occurs. However, many posts may not be expressive enough to provide the therapeutic power. In this study, we build a Gaussian naive Bayes model to detect expressive writing in an online health community, CaringBridge. Because we lack full text data as training data, we use a method to learn model parameters from meta-analysis of the literature. The classifier reaches reasonable accuracy on the test set annotated by the authors. We also explore factors that may influence users' spontaneous expressive writing. We find gender, health condition, author type, and privacy settings can affect individuals' spontaneous expressive writing. Finally, we reflect on our methodology and results and provide design implications for online health communities.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: online health community; CaringBridge; expressive writing; health blogging; review; meta-analysis; survey; classification; aggregated data; published result

ACM Reference Format:
Haiwei Ma, Sunny Parawala, and Svetlana Yarosh. 2021. Detecting Expressive Writing in Online Health Communities by Modeling Aggregated Empirical Data. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 62 (April 2021), 32 pages. https://doi.org/10.1145/3449136

1 INTRODUCTION

Expressive writing, which is writing about very deepest thoughts and feelings about a traumatic event, has been consistently proved to benefit writers' psychological well-being, physical health, and social behaviors [30, 70, 89]. The majority of participants in these studies reported that the writing experience was helpful and meaningful [68]. The *standard expressive writing* requires guidance by trained experts and has been mostly studied in the controlled lab environment [30]. This method ensures participants' compliance with expressive writing and has high internal validity. However, it suffers from poor external validity. Getting instructions from a writing therapist is not always possible in the real world. It's better that expressive writing and its therapeutic effects take place naturally in the self-administered writing.

Authors' addresses: Haiwei Ma, maxxx979@umn.edu, University of Minnesota, Minneapolis, Minnesota, USA, 55455; Sunny Parawala, paraw001@umn.edu, University of Minnesota, Minneapolis, Minnesota, USA, 55455; Svetlana Yarosh, lana@umn.edu, University of Minnesota, Minneapolis, Minnesota, USA, 55455.

2573-0142/2021/4-ART62 $15.00
https://doi.org/10.1145/3449136

With the development of Internet technologies, individuals can now post struggles along their health journeys on blogs, personal websites, or online health communities. Writing about health journeys will inevitably disclose individuals' deepest thoughts and feelings, so it is possible that expressive writing occurs spontaneously in this process [90]. Indeed, several studies showed that health blogging had some similar therapeutic effects as the standard expressive writing [43, 76]. Health blogging is also perceived by writers as helpful and meaningful, like the standard expressive writing [14, 94]. In this study, we refer to the natural occurrence of expressive writing in user-generated content like blogs as *spontaneous expressive writing* to differentiate it from the standard experimental method.

Nonetheless, self-disclosure (revealing information about oneself to others [25]) via blogs or other forms may not always be expressive enough to provide the health benefits [104]. For example, a post may only disclose factual details or about neutral topics (e.g. personal schedules or plans) without touching thoughts and feelings deep inside. People may be faced with self-presentational and privacy concerns that prevent them from disclosing deeper thoughts and feelings [5, 19, 99]. An automated analysis tool that predicts if a post is qualified as expressive writing will help individuals self-track their writing process to gain the most benefits from it. It can not only assist with individuals' self-therapy through writing, but also provide support for researchers to detect expressive writing from large amount of writings and study it outside labs. Unfortunately, we do not find any prior work that studied the prediction of expressive writing. Therefore, our first research question is:

RQ1: *How can we detect spontaneous expressive writing in user-generated content in the wild?*

In the standard experimental protocol, participants' compliance with expressive writing is monitored by researchers. However, whether individuals will write expressively is probably influenced by many factors in the real world, such as individual characteristics and social contexts. For example, people who are suffering may engage more with expressive writing than people who are satisfied with life, because the former group is in need of a way to vent out their feelings. Previous studies showed people who were stressed or sick indeed benefited more from expressive writing than people who were cheerful and healthy [30]. Studying potential factors that causes different writing behaviors (writing expressively vs. non-expressively) can help us understand how to encourage and facilitate expressive writing. To the best of our knowledge, there is no research looking into this aspect and thus, we propose our second research question:

RQ2: *What factors influence individuals' engagement with spontaneous expressive writing?*

The sensitive nature of expressive writing and the need to protect participant privacy prevent therapists and scientists from freely sharing the original essay datasets. We only have access to summaries of text data in aggregated forms in the publications (means and standard deviations of linguistic features measured for different writing conditions). Consequently, we cannot use the typical machine learning methods that construct models on individual data points. In order to address the first research question, we have to analyze the published aggregated data to train models. We first conducted a meta-analysis of empirical work that reported summaries of text data. It provides us with one of the best possible approximations of a ground truth for expressive writing (real linguistic features that characterize expressive writing), because it is based on summaries of linguistic features from studies that enforce rigorous procedures for expressive and non-expressive conditions, respectively. Given the available data and the specific characteristics of prior work on expressive writing, we chose Gaussian naive Bayes over other models. By assuming normal distribution and feature independence, Gaussian naive Bayes avoids computation of distribution and covariance that require individual data points [40]. For each feature, a normal distribution can be directly specified by its mean and standard deviation derived from meta-analysis.

The lack of full text datasets also poses a challenge for model evaluation. Unlike essays that are clearly separated into expressive vs. non-expressive conditions by the experimental manipulation, blog posts generated in the natural environment are more likely to be on a spectrum of expressiveness. There is no guidance on how to manually annotate expressive writing from less clear-cut essays in the literature. Therefore, we developed our own coding scheme from scratch and coded updates from a nonprofit online health community, CaringBridge, as test sets. In order to address the second research question, we applied the predictive model to all updates in CaringBridge and used a logistic regression model to study the relationship between different potential factors and individuals' engagement with spontaneous expressive writing.

In Section 2, we provide an overview of related works, which inspire our choices of features for classification model and factors that may influence individuals' engagement with spontaneous expressive writing. We go on to describe our research platform and methodology and explain the decisions we made in the process in Section 3. We then present results of classifier and regression in Section 4. Finally, we discuss these results and explain the limitations of our study in Section 5.

2 RELATED WORK

2.1 Expressive Writing

In the *standard expressive writing* protocol, participants are asked to write about their very deepest thoughts and feelings about traumatic experiences for 15-20 minutes on three to five laboratory sessions with no feedback given [68]. Since the first study conducted by Pennebaker [69], a number of subsequent studies have replicated this experimental procedure and found a broader range of beneficial effects of expressive writing on psychological well-being, physical health, and social behaviors [30, 70, 89]. The majority of participants in these studies reported that the writing experience was valuable and meaningful [68]. The beneficial effects of expressive writing were observed across different age groups[47, 97], genders [58], education levels [78, 91], ethnicity, languages, and cultures [26, 44, 56]. Patients of various health conditions and caregivers were also found to benefit from expressive writing [31, 61, 79].

The standard instructions were later modified to explore the boundary conditions of expressive writing and investigate the psychological mechanisms underlying it. Several studies showed that health benefits could still be found when participants were asked to focus on positive aspects of adverse experiences (benefit finding) [20], write in a narrative fashion about stressful events (narrative writing) [21], write gratitude letters [96], or even write about imaginary traumas [35]. These findings suggest that therapeutic effects will probably not be erased, as long as the alterations of standard protocols do not interrupt the evoking of emotional and cognitive processes during the writing [30, 51, 55, 57]. Indeed, emotional expression and cognitive restructuring are believed to be the key factors that contribute to the beneficial effects [30, 51, 57]. This belief is supported by the evidence that expressive writing increased working memory capacity and reduced intrusive and avoidant thinking (*cognitive restructuring*) and that expressive writing caused physiological activation and emotional arousal, which attenuated after multiple writing sessions (*disinhibition and habituation of emotions*) [46, 87].

Some studies also investigated home-based or Internet-based expressive writing. Participants were asked to conduct expressive writing at home or online with experimental instructions given in written and videotaped format [10, 81], or via web applications [17, 49]. Beneficial effects were still found in these expressive writing studies outside laboratory. Although these studies moved one step further from labs to the natural environment, they suffered from the poor external validity and required administration by researchers or therapists. In real-world settings, it is not always feasible or affordable to get supervision from therapists, which may create additional burdens on individuals. Therefore, it is valuable to explore *non-experimental, spontaneous expressive writing*, where individuals conduct expressive writing out of their own wills (not enforced by researchers) and in a natural way (not in the structured manner). Spontaneous expressive writing can be lower in cost, easier to access, and more enjoyable [86].

With the development in blogging technologies and online health communities, individuals can now disclose their thoughts and feelings about stressful experiences, especially their health journeys, in blog posts. A number of studies evidenced that health blogging had therapeutic power on mental health [14, 77, 94] and social health [43, 76]. Bloggers also reported that they found blogging activity to be helpful and engaging [14, 94]. Although the natural occurrence of expressive writing in health blogging provides great benefits, without enforcement of expressive writing instructions, many blog posts may only write about facts instead of personal thoughts and feelings. As a result, these writings fail to provide the therapeutic power [104]. To help individuals self-track their writing process and learn to write expressively on their own, we need an automated tool to tell whether an update is qualified as expressive writing. This tool also makes it possible for researchers to isolate spontaneous expressive writing from large amount of user-generated data and study expressive writing in the natural environment. To build such tool, we review the literature on language features that characterize expressive writing in Section 2.3.

2.2 Self-Disclosure

Self-disclosure refers to the process "by which one person verbally reveals information about himself or herself to another" [25]. People may have different motives for self-disclosure, which influence what and how much they decide to disclose [106]. For example, individuals will share positive personal experiences (positive self-disclosure) to maintain a positive self-image and impress others [32, 98]. They may also share negative personal experiences (negative self-disclosure) to solicit social support and help [106]. Self-disclosure has been investigated in a variety of health support groups on Reddit [27], Twitter [45], Cancer Survivor Network [106], Facebook, and Yahoo [2]. Several studies tried to build classifiers to detect positive and negative self-disclosure in online health support groups using a set of linguistic features [105, 106].

As one type of self-disclosure, expressive writing is special and interesting for several reasons. First, expressive writing is proved to have health benefits, while self-disclosure generally does not have therapeutic power itself. Second, expressive writing is longer, more sensitive, and includes more thoughts and feelings than self-disclosure in general (recall that participants are asked to write about deepest thoughts and feelings about traumatic experiences for 15-20 minutes). Third, expressive writing is less explored than self-disclosure in online health communities. Previous studies predicted self-disclosure on conversation-like posts and messages in forum threads or private channels using validated training data [106]. However, there is no research on the detection of expressive writing due to the lack of validated training data and the difficulty of labeling it. In this paper, we explore the possibility of using summaries of text data from empirical work as training data and predict expressive writing on diary-like CaringBridge updates.

2.3 Linguistic Markers of Psychological Processes

Linguistic markers has been used to quantify and detect various psychological processes. Researchers have designed algorithms using language features of posts on social media to predict whether an individual is suffering from depression [22], eating disorder [101], psychotic disorder [7], stress [16, 83], depression [102], and anxiety [85]. In addition, language use on social media can be used to infer mental health[13], personality [33], and mindfulness [15]. One text analysis approach that is commonly used in these studies is the **Linguistic Inquiry and Word Count (LIWC)** software, developed by Pennebaker and his colleagues as part of their work examining essays written in the expressive writing condition [72, 74]. Its lexicon includes a variety of linguistic categories which consist of words related to different psychometrics. The percentages of words from these different categories used in an essay can be computed.

The prior literature found that subjects in the expressive writing condition rated their essays as significantly more emotional, personal, and meaningful than controls [71]. Because expressive writing translates individuals' stressful experiences into words, researchers computed the percentage of emotion words (positive and negative emotion words in LIWC) and cognitive words (causal and insight words in LIWC) as linguistic clues to capture the emotional and cognitive processing behind writing [42]. It has been consistently found that use of emotion words and cognitive words are significantly higher in essays written in the expressive writing condition than those in essays written in the control condition [65, 80, 92]. In other words, expressive writing is characterized by higher percentage of emotion words and cognitive words compared to non-expressive writing. The difference in these LIWC scales is thus used by many experimental studies of expressive writing as manipulation check [64, 88].

As mentioned in Section 2.1, the key factors that lead to the therapeutic effects of expressive writing are believed to be cognitive restructuring and emotional expression [3]. If these psychological processes can be captured by LIWC features, researchers should be able to find a direct association between the word use computed by LIWC and improvements in health outcomes. Several studies investigated this possibility and found that higher use of positive emotion words (happy, laugh), negative emotion words (sad, angry), causal words (because, reason), and insight words (understand, realize) in essays were indeed associated with better health outcomes [12, 52, 53, 73, 77].

These evidence suggest that language features can capture the underlying emotional and cognitive processes in the writings and may be used to differentiate expressive writing from non-expressive writing. However, to the best of our knowledge, there is no work that use textual features to automatically classify expressive writing. In this study, we follow the traditional approach of using LIWC to obtain the percentage of emotion words and cognitive words and use these features to build a classifier that classifies whether a writing is expressive or not.

2.4 Explanatory Variables of Expressive Writing

We are also interested in what factors influence individuals' engagement with expressive writing, because it can help us understand how to encourage and facilitate expressive writing. There is no direct research on possible explanatory variables of expressive writing, so we reviewed the literature that studied predictors of self-disclosure or moderators of effects of expressive writing.

Gender. Prior work found that women disclosed more than men regardless of whether self-disclosure was measured by self-report or observation [24]. In online settings, females were also found to disclose loneliness more than males [45]. Psychologists showed that social norms influenced the different tendencies to disclose in men and women. A woman would be seen as better adjusted when she disclosed and a man would be seen as better adjusted when he did not [23]. Expressive writing is one type of self-disclosure. Therefore, we hypothesize: **H2a. Female writers are more likely to engage with expressive writing than male writers.**

Health Condition. Previous studies suggested that participants were more willing to disclose when they faced more severe problems [37], or perceived greater necessity of disclosure [1]. It was also found that participants who were stressed or sick received greater benefits from expressive writing than those who were not [30]. Participants who wrote about a more severe trauma received greater therapeutic effects than those who wrote about low-severity trauma [34, 66]. These evidences imply that people who battle life-threatening, or terminal diagnoses like cancer have greater desires to write and benefit more from it than people who face less-severe health conditions. Therefore, we hypothesize: **H2b. Writers who write about health journeys related to cancer are more**

likely to engage with expressive writing than writers who write about health journeys related to other health conditions.

Patient vs. Caregiver. Because many patients are battling life-threatening, or terminal diagnoses like cancer, their family caregivers may take the responsibility to write about health journeys for them. Patients are likely to endure more pain than caregivers through health journeys and thus, may have greater desires to vent out thoughts and feelings [30, 66]. Furthermore, caregivers mostly write from third-person perspectives and may use more objective tones than patient writers. Therefore, we hypothesize: **H2c. Writers who are patients are more likely to engage with expressive writing than writers who are family caregivers.**

Privacy. People may be unwilling to disclose due to privacy concerns [5, 99]. Without privacy protection, illness-related embarrassment or stigma may also appear [1, 75]. Prior work showed that individuals were more willing to disclose personal information, opinions, and feelings on anonymous sites [6, 60, 84]. It was also found that people disclosed more in private channels than public channels [106]. On CaringBridge, users are free to name their sites and can manage the audience by changing privacy settings from default medium (visible to all registered CaringBridge users except those whom they wish to block) to low (visible to anyone including non-registered visitors) or high (visible only to those whom are invited by the writer). A larger audience might cause greater self-presentational and privacy concerns. Therefore, we hypothesize: **H2d. Writers with more stringent privacy settings are more likely to engage with expressive writing than writers with lower privacy settings.**

3 METHODS

To address our research questions, we first conducted a meta-analysis of empirical work that reported the means and standard deviations of LIWC features of expressive writing. The overall means and standard deviations synthesized from meta-analysis were then used as model parameters that specify the normal distribution of each class in the Gaussian naive Bayes model. Next, we developed our own coding scheme to annotate CaringBridge updates as a test set and evaluated the performance of our classifier on it. The pipeline is shown in Figure 1. Before going into technical details, we first explain our research collaboration with CaringBridge and discuss ethics surrounding the relationship.

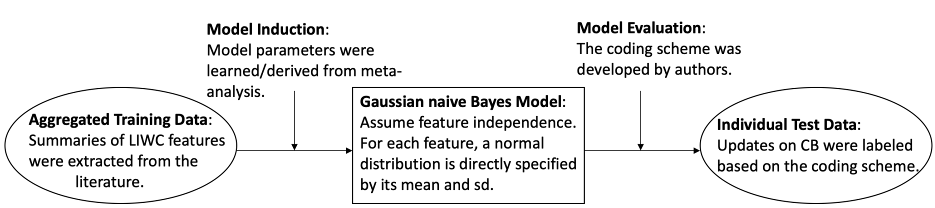

Fig. 1. The pipeline of building machine learning models to detect spontaneous expressive writing.

3.1 CaringBridge: Platform Description, Research Collaboration, and Ethics

CaringBridge is an online platform that connects patients and caregivers to family and friends through their health journeys. As a 501(c)(3) non-profit organization, CaringBridge has had over 2.8 billion cummulative visits from over 230 countries and territories around the world. The major difference between CaringBridge and other online health communities is its primary functionality, "journal", which resembles a blog and contains a collection of "updates" arranged in a timeline.

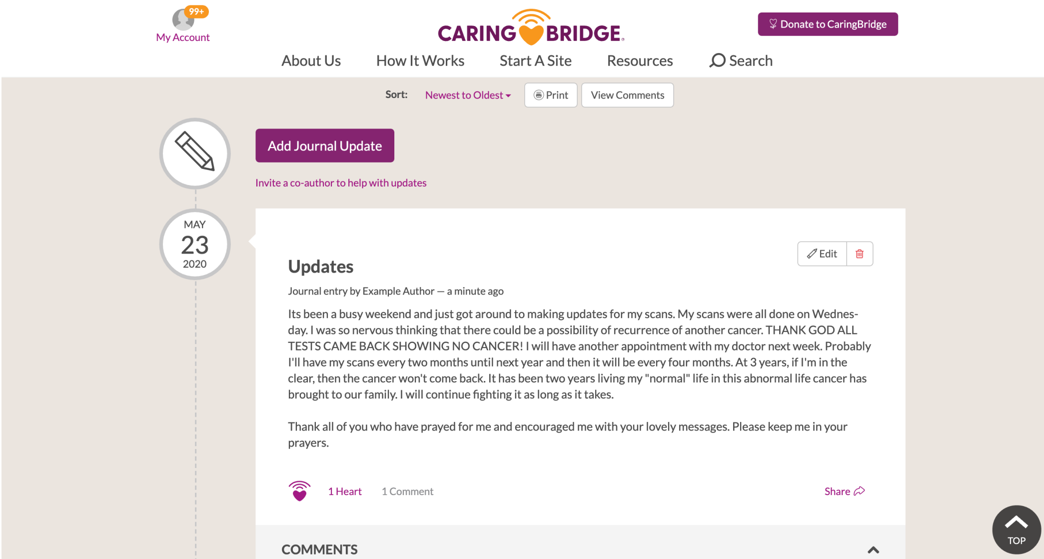

Fig. 2. An example site's homepage and journal. The journal update is modified from real user data.

CaringBridge users (patients, caregivers, or both) can start a "site" (similar to a personal website) to create a journal about a health journey and invite family and friends to follow and subscribe to it. Subscribers receive notifications when a new journal update is posted. A site can have multiple "authors" and an author can manage multiple sites. Users may optionally choose to disclose the nature of the health condition for their site. Cancer is most frequently reported health condition but other conditions include surgery/transplantation, injury, cardiovascular/stroke, COVID-19, etc. Authors can change the privacy settings to manage the audience, where they can choose from low (anyone can view the journal), medium (registered user of CaringBridge can view the journal, has a block list), or high privacy (approved visitors only, has an exclusive list of invitees). Figure 2 demonstrates the appearance of a site and a journal update.

This work proceeds from a research collaboration between CaringBridge and an interdisciplinary team from University of Minnesota. In accordance with CaringBridge's Privacy Policy & Terms of Use Agreement, the dataset includes de-identified information about 588,210 sites and 22,333,379 users between June 1, 2005 and June 3, 2016. Conducted with permission of the CaringBridge leadership, this study was reviewed and deemed exempt from further IRB review by the University of Minnesota Institutional Review Board. All text mining searches and computations performed in this study took place on the University of Minnesota's high performance computing clusters, which provided a secure location for the data and analysis.

As a collaborative endeavor, we find it important to introduce data use constraints to protect the privacy of individuals who use CaringBridge. CSCW and related fields have recently hosted a number of conversations to decide appropriate practices and norms around data use and dissemination. On one hand, it is critical to protect the rights of participants (above and beyond IRB review) [11] and no de-identification technique is without fail [63]. On the other hand, open data dissemination supports better science and forms of replication that could strengthen the field as a whole [38]. We respect both sides of this conversation. Given the sensitive nature of this data, it is not possible for us to publicly release the dataset used for analysis in this paper. However, we invite questions about the project and dataset by contacting the investigators who conducted this study or CaringBridge directly. We find this to be a reasonable compromise between the priorities of ethical protection and of replicable science.

3.2 Meta-analysis of Expressive Writing Literature

Due to its sensitive nature, there is no publicly released dataset of individual essays. Instead, we only have access to summaries of LIWC features in aggregated forms that characterize expressive writing. Most traditional machine learning methods require individual data points to learn model parameters and cannot deal with aggregated training data. For example, support vector machine needs to compute the dot product of input vectors [8]. One approach to utilize the aggregated data is meta-analysis. Meta-analysis refers to the systematic review of a set of individual studies with the aim to combine their results. It provides a better estimate of ground truth measurements than single studies due to the increased amount of data and statistical power [93]. The results of meta-analysis also generalize to a wider population due to the combination of different study populations [50]. It is thus a valuable tool to objectively summarize existing evidence regarding a specific issue [100]. In this paper, we used pooled estimates from meta-analysis as model parameters that summarize the normal distribution of each class in Gaussian naive Bayes and meta-analysis can be viewed as a learning process to determine model parameters using published aggregated data.

The meta-analysis of empirical work provides one of the best possible approximations of ground truth measurements for expressive writing and non-expressive writing, because: 1) summaries of LIWC features were extracted from experimental studies that enforced rigorous expressive writing and control conditions and often validated their design by manipulation check; 2) multiple studies were combined to reduce possible sampling error in single studies and increase the amount of data and statistical power; and 3) we applied random effects model for meta-analysis to account for the variance introduced by different study populations and increase the generalizability to a wider population. In contrast, annotation of CaringBridge updates to create individual training data will introduce the bias and reduce the data quality due to lack of validated guidance for labeling expressive writing and less-clear boundaries between expressive and non-expressive writing in the wild. Hence, the generated training data may not truly characterize expressive and non-expressive writing defined by domain experts. Furthermore, the meta-analysis includes about 1,100 essays to yield the final estimates of ground truths for positive and negative emotion words and about 500 essays for causal and insight words. To annotate a training dataset of similar size will take much time and efforts.

We started our search on Google Scholar database with keywords *"expressive writing" OR "disclosure writing" OR "written emotional expression" OR "written emotional disclosure" OR "therapeutic writing" OR "writing therapy"* to avoid missing any publications[1], because these terms are often used interchangeably in the literature. This returned 17,700 results, which exceeded our capability to process. Upon checking publications from the first few pages, we found that many of them only mentioned expressive writing in related work or references but did not really study it. Therefore, we added terms "control group" and "control condition" to restrict to those publications that studied expressive writing in experiment and had a control condition besides experimental conditions. We did not add terms "experimental group" or "experimental condition" because that would exclude many studies that used different names for their experimental groups (e.g. expressive writing group, disclosure group). We again checked publications from the first few pages and found most of them did not report LIWC word counts of essays. We then added the term "LIWC" to further limit to

[1]Note that the quotation marks around each phrase means finding the exact phrase.

those that mentioned LIWC. This returned about 658 results, which was a reasonable number. The final set of keywords are as follows:

> *("expressive writing" OR "disclosure writing" OR "written emotional expression" OR "written emotional disclosure" OR "therapeutic writing" OR "writing therapy") AND ("control group" OR "control condition") AND LIWC*

We employed a two-pass filtering process on the paper pool. In the first pass, we read the title and abstract to filter out papers that were not relevant to expressive writing and skimmed through the results to filter out papers that did not report LIWC summaries. This removed 563 papers and left us with 95 papers. In the second pass, we read the methods and results section and filtered out papers that did not enforce the standard expressive writing protocols or did not report means and standard deviations of word use in essays. Examples of excluded papers were papers that designed a modified version of standard protocols such as writing about positive events only (positive writing) [4], writing about autobiographical narratives (narrative writing) [28, 95], and online blogging [54], and papers that only reported differences of LIWC scores between conditions [67], or correlations of LIWC scores with some other measures [18]. This removed 48 papers and left us with 47 papers (see full list in Table 5 in appendix).

We are interested in the LIWC categories of positive emotion, negative emotion, causal, and insight, because previous studies already showed that higher use of these words characterized expressive writing and were associated with better health outcomes [65, 73]. Most studies only reported statistics of these categories. Only a few studies reported word use of other LIWC categories such as social (friend, family), first person pronouns (I, we, me), past tense verbs, present tense verbs, and future tense verbs [41, 103]. Since there is no evidence that these LIWC categories characterize expressive writing or predict therapeutic effects, we do not compute summary statistics for them.

We coded the remaining 47 papers and recorded the following values for each reported LIWC category we focused on: 1) mean percentage in the experimental condition (i.e., expressive writing condition); 2) standard deviation of percentage in the experimental condition; 3) number of participants' essays in the experimental condition; 4) mean percentage in the control condition (i.e., non-expressive writing); 5) standard deviation of percentage in the control condition; 6) number of participants' essays in the control condition. If the study tested several different experimental conditions, the one that resembled the standard expressive writing was picked. If the study included multiple writing sessions, we computed the average means and standard deviations across sessions. If the study reported standard error, the standard deviation was obtained from the standard error multiplying by the square root of the sample size.

We then applied random effects models to calculate the overall mean and standard deviation of percentage of positive emotion words, negative emotion words, causal words, and insight words in the experimental and control condition. Random effects models are common approaches for meta-analysis [9, 36]. Studies with greater precision (larger sample size and smaller standard deviation) are given a higher weight when the model combines results from different studies. Compared to fixed effects models, random effects models do not make the assumption that all included studies come from the same population [9, 36]. Instead, the studies included in the random effects meta-analysis are assumed to be a random sample of all possible studies. Therefore, it can account for the variance introduced by different study populations. Because studies that have missing values in mean, standard deviation, or sample size in the experimental or control condition cannot be used in the random effects meta-analysis for a given LIWC category, we ended up with 32 studies to derive overall means and standard deviations of positive emotion words in both conditions, 32 for negative emotion words, 15 for causal words, and 17 for insight words. The full list of papers included in the meta-analysis for each LIWC category is presented in Table 6, 7, 8, and 9, respectively in appendix. The results of meta-analysis are presented in Section 4.3.

3.3 Gaussian Naive Bayes Model

Gaussian naive Bayes model makes two assumptions: normality assumption (underlying data distribution being normal) and feature independence assumption (features being independent from each other). We chose Gaussian naive Bayes model over other models for several reasons. First, we only have summaries of LIWC features reported in the publications and do not have access to full text data (and would not be able to get such access to data from prior studies due to the personal nature of expressive writing and the need to protect participant privacy). Other models (e.g. support vector machine, random forest) require individual data points to train. Similarly, we need complete essays to extract other features (e.g. word embeddings) than the reported LIWC scores. Given the available data and the specific characteristics of prior work on expressive writing, Gaussian naive Bayes is a good choice. By assuming normal distribution and feature independence, we avoid computation of a full distribution and covariance that require individual data points. Second, previous studies demonstrated that naive Bayes models reached reasonable accuracy in text classification [59, 62]. For the novel task of classifying expressive writing without existing pre-trained models, Gaussian naive Bayes model based on prior experimental work provides a promising first step toward understanding expressive writing in the wild. We discuss the limitations of our model and important next steps to be able to move to more sophisticated features and models in Section 5.3.

Suppose we use random variable O to denote the percentage of positive emotion words, N to denote the percentage of negative emotion words, C to denote the percentage of causal words, I to denote the percentage of insight words, and U to denote whether expressive writing occurs in an update (1 means true, 0 means false). Each random variable (or feature) follows a normal distribution that differs between expressive and non-expressive writing (e.g.,$p(O|U = 0) \sim \mathcal{N}(\mu_{O0}, \sigma^2_{O0}), p(O|U = 1) \sim \mathcal{N}(\mu_{O1}, \sigma^2_{O1})$, since expressive and non-expressive writing are characterized by different language use. Thanks to the feature independence assumption, the joint distribution of these random variables is just the product of four normal distributions as follows:

$$\begin{aligned} p(O, N, C, I|U = 0) = p(O|U = 0)p(N|U = 0)p(C|U = 0)p(I|U = 0) \\ p(O, N, C, I|U = 1) = p(O|U = 1)p(N|U = 1)p(C|U = 1)p(I|U = 1) \end{aligned} \tag{1}$$

Based on Bayes rule, the posterior probability (the probability of an update being expressive given its language features) can be computed from the following:

$$p(U = 1|O, N, C, I) = \frac{p(O, N, C, I|U = 1)p(U = 1)}{p(O, N, C, I|U = 1)p(U = 1) + p(O, N, C, I|U = 0)p(U = 0)} \tag{2}$$

The probability of an update being non-expressive given its language features is just:

$$p(U = 0|O, N, C, I) = 1 - p(U = 1|O, N, C, I) \tag{3}$$

We did not have any prior knowledge of the relative frequency of expressive vs. non-expressive writing on CaringBridge, so we used uniform prior ($p(U = 0) = p(U = 1) = 0.5$). Using maximum likelihood estimation (MLE), unknown model parameters (e.g. μ_{O0} and σ^2_{O0} that specify the normal distribution followed by $p(O|U = 0)$) can be estimated by computing means and standard deviations of LIWC scores observed in the training data. Normally, we will calculate LIWC scores for each essay and then estimate their means and standard deviations from these individual data points. However, in our study, what we have is not individual data points, but a quantity of aggregated data from previous studies (i.e., means and standard deviations of LIWC scores summarized for a set of essays). We need a statistical tool to compute overall means and standard deviations for the entire combination of different sets of essays. Meta-analysis is a standard tool for this task and thus, we use its pooled estimates as model parameters. We also tested other priors (e.g. $p(U = 0) = 0.25, p(U = 1) = 0.75$) and did not find significant changes in the classifier's F1 score as shown in Table 3. This is evidence that the model classifies expressive writing mainly based on information in language features (likelihood function) rather than prior knowledge (prior distribution).

An update is then classified as expressive, if it has more than 300 words and its predicted probability of being expressive is greater than 0.5. Otherwise it is classified as non-expressive. We decided to choose 300 words as threshold. First, standard expressive writing protocols require individuals to write about 15 minutes and almost all studies in the literature reported more than 300 words in the expressive writing condition. It's very unlikely that expressive writing occurs if individuals write briefly. Therefore, we chose 300 words to approximate 15-20 minutes writing in the standard expressive writing. Second, percentages of words from different categories are unreliable when the update is very short. Just using one positive word will result in a high percentage if the update only has a few words.

3.4 Evaluation of Classifier Performance

The lack of full text datasets also poses a challenge for model evaluation. Although the model is trained on aggregated training data, it needs to classify each individual update, so we want to know its performance on individual test data. In addition, since model parameters are learned from essays written in controlled experiment, we want to know how well it generalizes to essays written in the wild, a test set which is completely unknown to the model a priori and originates from different sources. Because there is no guidance on how to label expressive writing, we need to perform a content analysis and develop our own codebook to annotate test data [39]. CaringBridge journal updates with fewer than or equal to 300 words do not meet the standard of 15 minutes writing required by expressive writing and thus, should be classified as non-expressive. Inclusion of these updates will not provide any insights on the model's real performance, but only imbalance test data. We excluded 9,782,929 updates with fewer than or equal to 300 words and retained 3,681,174 updates with more than 300 words for further content analysis.

We decided to analyze each writing at the sentence level, because a sentence usually contains one message and provides a moderate level of granularity for judgement. A word or phrase is too short to convey a complete message, whereas an update is a mix of multiple messages. It is thus more difficult to judge the content and develop coding criteria at the word or update level. We independently conducted an exploratory reading of 100 randomly selected updates. In this process, we first distinguished sentences about thoughts and feelings from sentences about facts, because based on its definition, writing about thoughts and feelings is a necessary condition for expressive writing (i.e. sentences signaling expressive writing must be about thoughts and feelings). However, not all sentences about thoughts and feelings signal expressive writing. Among these sentences, we further separated "key" sentences that clearly signaled expressive writing from ambiguous ones that were less clear on whether they signaled expressive writing and marked them in different colors. In this way, we can formulate other necessary conditions for expressive writing and develop a codebook of higher validity. Updates with at least one key sentence were labeled as expressive writing and updates without any key sentence were labeled as non-expressive writing.

We then discussed each key sentence and each ambiguous sentence to clarify the boundary of expressive writing vs. non-expressive writing. We identified another two necessary conditions based on analysis of those sentences and definition of expressive writing. First, sentences signaling expressive writing must be writing about self. Sentences about others' thoughts and feelings and sentences quoted from books or other sources are not considered as expressive writing, because they are not self-disclosure. Second, sentences signaling expressive writing must be writing about health journeys or other life struggles. Sentences about mundane activities, social events, or everyday objects and sentences about expressions of gratitude, requests of various support, or sending of good wishes, even though they may touch individuals' own thoughts and feelings, are not considered as expressive writing, because they are not about deep topics. Although none of the three conditions are sufficient, a combination of them should be close to the necessary and sufficient condition of expressive writing. In other words, a sentence satisfying all three criteria (about thoughts and feelings, about self-disclosure, and about specific or general struggles) is likely to imply expressive writing. Conversely, a sentence that does not meet all three rules is unlikely to indicate expressive writing. The refined codebook to label expressive writing is as follows:

- An update is labeled as expressive writing, if it contains at least one key sentence that clearly signals expressive writing. An update is labeled as non-expressive writing, if it does not have any key sentence.
- A sentence is a key sentence if it discloses: 1) one's general feelings, moods, or mental states that are not directed at, related to, or preceded by a specific event or object; 2) one's thoughts and emotions about stressful events; 3) one's reflection on the health journey (e.g. what he/she has learned, what he/she would like to change); or 4) one's coping with struggles and difficulties (e.g. discovery of meaning, accepting life, taking a self-distanced perspective, believing and praising God).
- A sentence is not a key sentence if it writes about: 1) facts, descriptions, plans, or schedule; 2) one's thoughts and emotions about mundane activities, social events, or everyday objects (e.g. we visited some old friends and felt happy); 3) others' thoughts and emotions (e.g. I know she must be upset); 4) content quoted from books, bibles, or other sources; or 5) common expressions in online health communities (e.g. requesting prayers, messages, visits, or other support, expressing gratitude, sending good wishes).

We then independently labeled another 100 updates to compute inter-rater reliability (coding instructions are in appendix F). The Cohen's Kappa of two raters is 0.64, which is acceptable but not very good due to the difficulty and complexity of this labeling task. Unlike many other labeling tasks, boundary conditions of expressive writing are not well understood and thus, there lacks guidance for labeling expressive writing. Besides, the borders between expressive writing and non-expressive writing are less clear in the wild than in the laboratories. There is also subjectivity in interpretations of thoughts and feelings, self-disclosure, and life struggles. Despite this limitation, we do not have better options for model evaluation. We went through this second set of 100 updates together and resolved any disagreements. We ended up with 200 updates whose labels are discussed and agreed by both raters. The results of model performance on this test set are presented in Section 4.3.

3.5 Explanatory Variables of Spontaneous Expressive Writing

As discussed in Section 2.4, we explored the following independent variables that might influence spontaneous expressive writing:

- **Gender:** We use a binary variable to denote whether an update is written by female (1) or male (0) writers.
- **Health Condition:** We combine different kinds of cancer into one group and other health conditions (e.g. surgery/transplantation, injury, cardiovascular/stroke) into another group. We use a binary variable to denote whether an update is written about health journeys of cancer (1) or other health conditions (0).

- **Patient vs. Caregiver:** We use a binary variable to denote whether an update is written by patient (1) or caregiver (0) writers.
- **Privacy:** We use a categorical variable to denote whether an update is posted under low (0), medium (1), and high (2) privacy. On CaringBridge, low privacy means information is visible to anyone including non-registered visitors. Medium privacy means information is visible to all registered CaringBridge users except being blocked. High privacy means information is only visible to those whom are invited by the writer.

These independent variables can be directly mined from CaringBridge datasets of user profiles and site settings. The dependent variable is the predicted label of whether an update is expressive or not. We used logistic regression to test the relationship between predictors and the response variable. The results are in Section 4.4.

4 RESULTS

4.1 Descriptive Statistics

There are about 13,464,103 journal updates in CaringBridge data. Table 1a reports the descriptive statistics of LIWC features and total words. The correlations among them are mostly less than 0.1, with a few exceptions noted in Table 1b. The relatively low correlations suggest our feature independence assumption made in naive Bayes model is indeed reasonable.

%	Mean	Median	SD
posemo	0.051	0.047	0.032
negemo	0.013	0.010	0.014
causal	0.011	0.010	0.011
insight	0.017	0.015	0.014
length	243	171	251.2

(a) Descriptive statistics

Variable Pair	Corr.
(posemo, negemo)	-0.10
(posemo, length)	-0.14
(causal, length)	0.11
(causal, insight)	0.11

(b) Correlation

Table 1. (a) Descriptive statistics of percentage of words from different LIWC categories (posemo for positive emotion, negemo for negative emotion) and count of total words (length). (b) Correlation between features larger than 0.1 or smaller than -0.1.

4.2 Meta-Analysis

We used meta-analysis to learn model parameters from aggregated data in publications. Results in Table 2 show the results of meta-analysis. Essays written in the expressive writing condition on average had 2.61% of total words being positive emotion words (sd=0.69%), 2.68% of total words

	Experimental		Control		Independent t test	
LIWC category	**mean(%)**	**sd(%)**	**mean(%)**	**sd(%)**	$\|t\|$	p**-value**
positive emotion	2.61	0.69	1.67	0.58	6.28	<0.001***
negative emotion	2.68	0.74	0.74	0.56	14.84	<0.001***
causal	1.64	0.66	0.83	0.41	6.65	<0.001***
insight	3.00	0.67	1.05	0.57	8.48	<0.001***

Table 2. Results of random effects meta-analysis of expressive writing literature on LIWC features.

being negative emotion words (sd=0.74%), 1.64% of total words being causal words (sd=0.66%), and 3.00% of total words being insight words (sd=0.67%). We find that people use more negative emotion words and insight words than positive emotion words in the expressive writing condition. This is evidence that thoughts and feelings deep inside are disclosed in the experimental condition. Essays written in the control condition on average had 1.67% of total words being positive emotion words (sd=0.58%), 0.74% of total words being negative emotion words (sd=0.56%), 0.83% of total words being causal words (sd=0.41%), and 1.05% of total words being insight words (sd=0.57%). We find that people use more positive emotion words than other categories in the non-expressive writing condition. This is evidence that thoughts and feelings deep inside are not touched in the control condition.

We also find that essays written in the experimental condition (expressive writing) have significantly higher percentages of positive emotion words, negative emotion words, causal words, and insight words than those in the control condition (non-expressive writing). The mean percentage of negative emotion words and insight words in expressive writing is about three times the mean in non-expressive writing. It again confirms the feasibility of using LIWC features to differentiate expressive writing from non-expressive writing, as shown by the prior work [65, 80, 92]. The pooled means and standard deviations of LIWC features are used as model parameters of Gaussian naive Bayes classifier.

4.3 Classifier for Expressive Writing

Our classification model is built upon parameters pooled from meta-analysis. We evaluated our algorithm on the test set of 200 updates with over 300 words. Table 3 presents the results of our classifier. The accuracy of our classifier is 67% with uninformative prior ($p(U = 0) = p(U = 1) = 0.5$). This result addresses our first research question: automatic classification of expressive writing in online updates is feasible without using a full training data. We can employ meta-analysis to pool means and standard deviations from summaries of LIWC features reported in papers and use Gaussian naive Bayes for classification, because the pooled means and standard deviations can be directly used as model parameters. As the first attempt to classify expressive writing, this accuracy is reasonable given that we lack full dataset to extract more sophisticated features and train more advanced models and that the model evaluation is on CaringBridge updates, which are completely unknown to the model a priori and originate from different sources.

We find that the model has high recall and low precision. This suggests our model is more likely to mistakenly classify a non-expressive update as expressive than to mistakenly classify an expressive update as non-expressive. This is probably because we carried out a strict coding scheme to annotate the test set. Recall that none of writing about one's thoughts and emotions about mundane activities, social events, or everyday objects, about others' thoughts and emotions, about content quoted from books, bibles, or other sources, or about common expressions to request support and show gratitude are considered as expressive writing by human raters. Such writing involves the use of emotional and cognitive words and is likely to be mistakenly classified as expressive writing by the algorithm. In the natural environment, there is no writing instructions or administrations from researchers, so greater noise and variance in the content is inevitable. Even human raters have low Cohen's Kappa (0.64) when labeling this test set.

We also test how different priors affect the model performance. We find that F1 score remains almost the same. Models with different priors all perform significantly better than the baseline random model. Increasing the prior probability of an update being expressive from 0.1 to 0.9 means the model is more likely to predict an update to be expressive and thus, has higher recall and lower precision. The model has highest accuracy when the prior distribution assumed ($p(U = 1) = 0.25$, 25% of updates are expressive) is closest to the actual relative class frequency in the test data.

Classifier	Performance			
	ACC	PREC	REC	F1
Random Model	0.49	0.52	0.5	0.51
Gaussian naive Bayes ($p(U = 1) = 0.5$)	0.67	0.63	0.92	0.75
Gaussian naive Bayes ($p(U = 1) = 0.1$)	0.66	0.63	0.86	0.73
Gaussian naive Bayes ($p(U = 1) = 0.25$)	0.68	0.64	0.90	0.75
Gaussian naive Bayes ($p(U = 1) = 0.75$)	0.64	0.60	0.94	0.73
Gaussian naive Bayes ($p(U = 1) = 0.9$)	0.61	0.58	0.97	0.73

Table 3. Model Performance evaluated on a test set of 200 updates. Models include: Random model that randomly guesses the label of an update; and Gaussian naive Bayes models with different priors. The model with uninformative priors ($p(U = 1) = 0.5$) is our final model. Column headers are average accuracy (ACC), precision (PREC), recall (REC), and f1 score (F1).

	Coefficients	Standard Error	p value
Intercept	-0.68	0.02	<0.001***
Gender: female (vs. male)	0.17	0.01	<0.001***
Health Condition: cancer (vs. non-cancer)	0.17	0.01	<0.001***
Author Type: patient (vs. caregiver)	0.23	0.01	<0.001***
Privacy: medium (vs. low)	-0.77	0.02	<0.001***
Privacy: high (vs. low)	-0.91	0.03	<0.001***

Table 4. Results of the logistic regression model testing effects of explanatory variables on whether an individual writes the update expressively or not. ***:p<0.001, **:p<0.01, *:p<0.05. The generalized variance inflation factor (GVIF) is computed for each variable: GVIF(Gender)=1.00, GVIF(Health Condition)=1.06, GVIF(Author Type)=1.06, GVIF(Privacy)=1.01. Updates with missing values in any of the independent variables are excluded, resulting in 464,552 observations in the logistic regression.

When we do not have any prior knowledge of the domain, uninformative prior should be used ($p(U = 1) = 0.5$), because it will avoid the situation where the assumed prior distribution is far from the actual frequency distribution. However, if prior knowledge of the domain is available, more informative priors may be used to improve accuracy. In our case, we do not have any prior knowledge of how many CaringBridge updates are expressive and how many are not, so we choose uninformative prior for our model. We apply it on the entire dataset and find that 2,978,293 updates are predicted as expressive (22%) and 10,485,810 (78%) are predicted as non-expressive.

4.4 Logistic Regression Results

Results in Table 4 show the relationship between different explanatory factors and users' spontaneous expressive writing. Updates with missing values in any of the independent variables are excluded, resulting in 464,552 observations in the logistic regression. We test multicollinearity of the regression model by computing generalized variance inflation factor (GVIF) for each independent variable [29]. We do not find any collinearity among explanatory variables as all GVIF values are relatively small (listed in Table 4). We find that all independent variables show significance in the model.

Specifically, the log odds of writing expressively in an update increases by 0.17 from male writers to female writers. This supports H2a that females writers are more likely to engage with expressive writing than male writers. The log odds of writing expressively in an update increases by 0.17 from writers who write about health journeys of other health conditions to writers who write about health journeys of cancer. This supports H2b that writers who write about health journeys of cancer are more likely to engage with expressive writing than writers who write about health journeys of other health conditions. The log odds of writing expressively in an update increases by 0.23 from caregiver writers to patient writers. This supports H2c that patient writers are more likely to engage with expressive writing than caregiver writers. The log odds of writing expressively in an update decreases by 0.77 from writers with low privacy settings to writers with medium privacy settings and decreases by 0.91 from writers with low privacy settings to writers with high privacy settings. This rejects H2d. Actually writers with higher privacy settings are less likely to engage with expressive writing than writers with lower privacy settings.

These results address our second research question: spontaneous expressive writing is indeed affected by a variety of factors, as implied in the literature (see Section 4.3). Most of our hypotheses are supported, but effects of privacy settings on users' engagement with expressive writing are opposite to our hypothesis. One possible explanation is that writers with higher privacy settings are those who tend to worry more about privacy or self-presentational concerns in nature or are experiencing illness-related embarrassment and stigma, which prevent them from expressive writing. Another possible explanation is that the results are confounded by potential benefits of expressive writing. Lower privacy settings may help people get more help, support, and responses from the community and give them a sense of being heard. It also avoids the inconvenience that family and friends have to register and log in CaringBridge to follow and view updates. Based on cost-benefit analysis, if people gain more benefits from lower privacy settings, they will still choose it despite higher risks. Furthermore, since people are free to name their sites on CaringBridge, lower privacy settings may not actually lead to higher perceived risks.

5 DISCUSSION

5.1 Reflection on Methodology

Classification of expressive writing is a novel problem that has not been explored. Its biggest challenge is lack of full text dataset. Because expressive writing involves disclosure of sensitive content, previous studies only reported summaries of LIWC features in the form of means and standard deviations without sharing the dataset. Traditional machine learning models require individual data points to train. In order to utilize the aggregated data, we conducted a meta-analysis to combine the results of different studies and derive overall means and standard deviations as model parameters used by Gaussian naive Bayes. Similar methodology can be adopted in other application domains, such as health care, where individual patient data is not accessible and only summaries of patient data are reported to protect privacy.

Compared to creation of training data from scratch for expressive writing, using existing aggregated data in the literature has several advantages. First, summaries of text data in the literature are collected and validated by domain experts. We can have confidence that the derived model parameters from meta-analysis truly characterize expressive writing. If we annotate training data on our own, it is likely to introduce bias and reduce data quality, because there does not exist any guidance on how to label expressive writing. Second, meta-analysis combines multiple studies and thus, includes a large size of samples to compute the final estimates (about 1000 essays to compute estimates for positive and negative emotion words). In contrast, annotation takes much time and efforts. It is very unlikely to create a dataset of similar size as meta-analysis. Furthermore, annotation itself can be a difficult task, because there are greater noise and variance in the content generated in the wild and boundaries between expressive and non-expressive writing on such content are fuzzy. Indeed, Cohen's Kappa between two human raters is not very high. Nonetheless, if there is guidance to annotate and validate data, it will be worth the pain to create a gold standard dataset that can be reusable by oneself and others and enable more sophisticated features and models.

5.2 Reflection on Results

Expressive writing can provide many health benefits, but it has been mostly studied inside laboratories and under experimental procedures. Expressive writing that occurs naturally outside the labs is rarely studied, partially because we are overwhelmed by large amount of writings and it's hard to find out the ones that are expressive and therapeutic. In this paper, we made the first step to differentiate expressive vs. non-expressive writing. By applying the classifier that classifies expressive essays, research in this space can be expanded from lab-based experimental studies to investigation of natural occurrences of expressive writing. Research in spontaneous expressive writing will benefit real users at a larger scale. In addition, this classification model may be applied to assist individuals self-track their writing process and reflect on it. In this way, users can gain the most benefits from their writing and learn to cope with the hard times in their life through expressive writing.

Among the 13M updates on CaringBridge, roughly 22% of them are predicted to be expressive. This suggests CaringBridge provides a space for many users to engage with expressive writing and gain benefits from it. From the logistic regression results, our hypotheses that female writers (H2a), writers who write about health journeys of cancer (H2b), and patient writers (H2c) are more likely to engage with expressive writing. For these users, it's important to provide an enjoyable experience of writing and keep their engagement so that they can receive health benefits. However, special care and support might also be needed, as such spontaneous expressive writing might indicate writers are suffering from greater pain and stress than others, which causes their greater motives to vent out thoughts and feelings and greater needs to solicit help from the community. For users who are less likely to engage with expressive writing, using prompts or incentives may facilitate their engagement with expressive writing, but too much intervention should be avoided. It should remain a free space for an individual to decide whatever he/she wants to write or decide not to write at all.

Contrary to our hypothesis H2d, we see that individuals are more likely to engage with expressive writing in lower privacy settings than in higher privacy settings. This implies that the existence of a larger audience is probably an incentive for writers to share their personal feelings and thoughts. Expressive writing is probably more than an individual activity, where people talk to themselves and heal mental wounds alone. Instead, expressive writing may serve for social purposes. It is possible that individuals write expressively in order to be heard and receive social support. It is also possible that expressive writing, like spiritual fellowship, is an important way to connect writers and readers together and strengthen their relations. However, it must be noted that in some contexts, expressive writing might harm social relations (e.g. when writing about something that might bring stigma or embarrassment) and high levels of privacy are still needed to create a safe environment for these individuals (such as 12 step followship [82, 107]).

5.3 Limitations and Future Work

By using Gaussian naive Bayes model, we make the assumptions that word use of different LIWC categories are normally distributed and language features are independent from each other. However, these assumptions may not be true. It is possible that percentages of words from different LIWC categories follow other distributions and different features are not independent. Our model only provides a reasonable approximation of the real data. However, these two assumptions are probably the minimal set of assumptions we have to make. We assume normal distribution, because we do not have full datasets to compute the exact distribution or prior knowledge of underlying data distribution. Similarly, we assume feature independence, because we do not have individual data points to compute the covariance between features. In the future, classification models that assume other distributions than Gaussian (e.g. beta distribution) may be explored.

In this work, we try to predict spontaneous expressive writing in online health communities that share similar linguistic characteristics with standard expressive writing in lab-based settings, because therapeutic effects and LIWC features of standard expressive writing are well documented. Given the diversity of content written in the natural environment, it is possible that there exist other types of spontaneous expressive writing, which has similar health benefits, but has different linguistic features. Psychologists have made efforts to push the boundary conditions of expressive writing and explore underlying mechanisms for its therapeutic power. For example, several studies showed that health benefits were also observed when participants wrote about positive aspects of adverse experiences (benefit finding) [20], wrote in a narrative fashion about stressful events (narrative writing) [21], wrote gratitude letters [96], or even wrote about imaginary traumas [35]. On CaringBridge, users also express gratitude or quote stories from books and bibles in their updates. However, evidence for beneficial effects of such "non-standard" expressive writing and reports of their linguistic characteristics are relatively scarce. We chose to follow the standard concept of expressive writing, because it has been well studied and established. The cost is that our model is not able to capture "non-standard" expressive writing. Future work can work on capturing all kinds of expressive writing or even predicting a continuous expressiveness score, if health benefits are consistently found for "non-standard" expressive writing.

We evaluated the performance of our classifier on a test set of CaringBridge updates annotated by ourselves. This will inevitably introduce bias into the data, since there is no validated annotation instructions. Different people may interpret the same sentence or writing differently, as there is no clear definition of what count as "deepest thoughts and feelings". Indeed, when we annotate the CaringBridge updates, Cohen's Kappa is just 0.64. Nonetheless, this is the best available method to test our classifier's performance. Our model is trained on aggregated data from essays in lab studies rather than CaringBridge updates, which means the test data is completely unknown to the model a priori and originates from different sources. Hence, we are unlikely to have overfitting issues. In the future, it will be useful to construct a validated approach to annotate expressive writing and compare the performance of model parameters learned from meta-analysis to those learned from grounded data.

We did not reach a high accuracy on the test set, but it is acceptable considering: 1) this is the first work to classify expressive writing and thus, neither prior knowledge nor pre-trained models can be utilized; 2) there lacks full dataset of individual data points to extract more sophisticated features or train more advanced models; 3) most previous studies did not report summaries of other linguistic features than emotional and cognitive words; 4) evaluation on a completely unknown test set extracted from a totally different source naturally leads to lower accuracy; 5) our annotation of test set introduces bias and leads to lower accuracy (e.g. disclosing one's thoughts and feelings about daily activities is not considered as expressive writing by us, but is classified as expressive writing by the model due to its use of emotional and cognitive words). Future work may focus on creation of gold standard datasets to enable the use of more sophisticated features and machine learning algorithms.

There are several directions for classifiers' improvement depending on the type of data to use. The first is to keep using aggregated data. We can try out other generative models with different assumptions regarding data distribution and evaluate their performance. We can also try to obtain other features in aggregated form than LIWC categories (e.g. term frequency–inverse document frequency) by collaborating with psychologists in this space and providing them with tools to compute those features. With more proper assumptions and more features, model performance can be improved without compromising participants' privacy. The second is to create a gold standard dataset for expressive writing. We may replicate previous lab-based studies and collect participants' essays in different writing conditions. These essays can then be directly used without additional labeling, as the condition determines the label. We may also construct a dataset from publicly available data such as posts in online health communities. This requires iterative codebook development in collaboration with domain experts to make sure that labels are reliable and valid. Future research may examine the necessary and sufficient condition for expressive writing, which will guide the codebook development. Future work may also look at the frequency rather than the existence of key sentences, or analyze content at a different level of granularity to develop a better codebook. With a full text dataset that enables all kinds of features and models, classifiers' performance can be improved. However, we must deliberately take steps to protect users' privacy, such as deidentifying users and storing data on secure servers. The third is to use existing datasets that are related to expressive writing. For example, datasets of self-disclosure classification (self-disclosure vs. not self-disclosure), topic classification (mundane, shallow topic vs. serious, deep topic), or subjectivity classification (opinions and attitudes vs. facts and descriptions) may be used for transfer learning to improve model performance [48], as expressive writing is related to these domains. The downside is its technical complexity.

6 CONCLUSION

This paper offers two major contributions to the CSCW community. First, we learn model parameters from meta-analysis of the literature to deal with the lack of full text data for our classification model and reach reasonable accuracy in predicting if an update is expressive or not. The methodology itself provides a new possibility for researchers to build automated tools for sensitive data that is only shared in summaries rather than in individual data points. The classification model built is useful to individuals in self-tracking of the writing process and researchers in isolating expressive writing from large amount of data. Second, we explore potential factors that may influence individuals' spontaneous expressive writing and identify gender, health condition, author type, and privacy settings as important factors. This serves as the first step to analyze expressive writing in the natural environment.

REFERENCES

[1] Robert R Agne, Teresa L Thompson, and Louis P Cusella. 2000. Stigma in the line of face: Self-disclosure of patients' HIV status to health care providers. (2000).

[2] Tawfiq Ammari, Sarita Yardi Schoenebeck, and Meredith Ringel Morris. 2014. Accessing social support and overcoming judgment on social media among parents of children with special needs. In *Eighth International AAAI Conference on Weblogs and Social Media*.

[3] Karen A Baikie and Kay Wilhelm. 2005. Emotional and physical health benefits of expressive writing. *Advances in psychiatric treatment* 11, 5 (2005), 338–346.

[4] Jonathan Britten Banks et al. 2004. The Impact of Event Familiarity on the Complexity and Coherence of Children's Narratives of Positive Events. (2004).

[5] Gaurav Bansal, David Gefen, et al. 2010. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision support systems* 49, 2 (2010), 138–150.

[6] Michael S Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Gregory G Vargas. 2011. 4chan and/b: An Analysis of Anonymity and Ephemerality in a Large Online Community.. In *ICWSM*. 50–57.

[7] Michael Leo Birnbaum, Sindhu Kiranmai Ernala, AF Rizvi, Elizabeth Arenare, AR Van Meter, M De Choudhury, and John M Kane. 2019. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. *NPJ schizophrenia* 5, 1 (2019), 1–9.

[8] Christopher M Bishop. 2006. *Pattern recognition and machine learning*. springer.

[9] Michael Borenstein, Larry V Hedges, Julian PT Higgins, and Hannah R Rothstein. 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods* 1, 2 (2010), 97–111.

[10] Joan E Broderick, Arthur A Stone, Joshua M Smyth, and Alan T Kaell. 2004. The feasibility and effectiveness of an expressive writing intervention for rheumatoid arthritis via home-based videotaped instructions. *Annals of Behavioral Medicine* 27, 1 (2004), 50–59.

[11] Amy S Bruckman, Casey Fiesler, Jeff Hancock, and Cosmin Munteanu. 2017. CSCW research ethics town hall: Working towards community norms. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 113–115.

[12] R Sherlock Campbell and James W Pennebaker. 2003. The secret life of pronouns: Flexibility in writing style and physical health. *Psychological science* 14, 1 (2003), 60–65.

[13] Stevie Chancellor, Eric PS Baumer, and Munmun De Choudhury. 2019. Who is the" Human" in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–32.

[14] Deborah S Chung and Sujin Kim. 2008. Blogging activity among cancer patients and their companions: Uses, gratifications, and predictors of outcomes. *Journal of the American Society for Information Science and Technology* 59, 2 (2008), 297–306.

[15] Susan E Collins, Neharika Chawla, Sharon H Hsu, Joel Grow, Jacqueline M Otto, and G Alan Marlatt. 2009. Language-based measures of mindfulness: Initial validity and clinical utility. *Psychology of Addictive Behaviors* 23, 4 (2009), 743.

[16] Glen Coppersmith, Craig Harman, and Mark Dredze. 2014. Measuring post traumatic stress disorder in Twitter. In *Eighth international AAAI conference on weblogs and social media*.

[17] Joanna Grace Crawford. 2019. WRITING FOR HEALTH: DEVELOPMENT AND EVALUATION OF AN INTERNET-BASED BENEFIT-FINDING WRITING INTERVENTION FOR ADULTS WITH TYPE 1 OR TYPE 2 DIABETES. (2019).

[18] J David Creswell, Suman Lam, Annette L Stanton, Shelley E Taylor, Julienne E Bower, and David K Sherman. 2007. Does self-affirmation, cognitive processing, or discovery of meaning explain cancer-related health benefits of expressive writing? *Personality and Social Psychology Bulletin* 33, 2 (2007), 238–250.

[19] S Nicole Culos-Reed, Lawrence R Brawley, Kathleen A Martin, and Mark R Leary. 2002. Self-presentation concerns and health behaviors among cosmetic surgery patients. *Journal of Applied Social Psychology* 32, 3 (2002), 560–569.

[20] Sharon Danoff-Burg, John D Agee, Norman R Romanoff, Joel M Kremer, and James M Strosberg. 2006. Benefit finding and expressive writing in adults with lupus or rheumatoid arthritis. *Psychology and Health* 21, 5 (2006), 651–665.

[21] Sharon Danoff-Burg, Catherine E Mosher, Asani H Seawell, and John D Agee. 2010. Does narrative writing instruction enhance the benefits of expressive writing? *Anxiety, Stress, & Coping* 23, 3 (2010), 341–352.

[22] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *Seventh international AAAI conference on weblogs and social media*.

[23] Valerian J Derlega and Alan L Chaikin. 1976. Norms affecting self-disclosure in men and women. *Journal of Consulting and Clinical Psychology* 44, 3 (1976), 376.

[24] Kathryn Dindia and Mike Allen. 1992. Sex differences in self-disclosure: A meta-analysis. *Psychological bulletin* 112, 1 (1992), 106.

[25] Kathryn Dindia, M Allen, R Preiss, B Gayle, and N Burrell. 2002. Self-disclosure research: Knowledge through meta-analysis. *Interpersonal communication research: Advances through meta-analysis* (2002), 169–185.

[26] Benjamin Dominguez, Pablo Valderrama, María de los Angeles Meza, Sara Lidia Pérez, Amparo Silva, Gloria Martínez, Victor Manuel Méndez, and Yolanda Olvera. 1995. The roles of disclosure and emotional reversal in clinical practice. (1995).

[27] Jordan Eschler, Zakariya Dehlawi, and Wanda Pratt. 2015. Self-Characterized Illness Phase and Information Needs of Participants in an Online Cancer Forum.. In *ICWSM*. 101–109.

[28] Robyn Fivush, Valerie J Edwards, and Jean Mennuti-Washburn. 2003. Narratives of 9/11: Relations among personal involvement, narrative content and memory of the emotional impact over time. *Applied Cognitive Psychology: the Official Journal of the Society for Applied Research in Memory and Cognition* 17, 9 (2003), 1099–1111.

[29] John Fox and Georges Monette. 1992. Generalized collinearity diagnostics. *J. Amer. Statist. Assoc.* 87, 417 (1992), 178–183.

[30] Joanne Frattaroli. 2006. Experimental disclosure and its moderators: a meta-analysis. *Psychological bulletin* 132, 6 (2006), 823.

[31] Pasquale G Frisina, Joan C Borod, and Stephen J Lepore. 2004. A meta-analysis of the effects of written emotional disclosure on the health outcomes of clinical populations. *The Journal of nervous and mental disease* 192, 9 (2004), 629–634.

[32] Jennifer L Gibbs, Nicole B Ellison, and Rebecca D Heino. 2006. Self-presentation in online personals: The role of anticipated future interaction, self-disclosure, and perceived success in Internet dating. *Communication research* 33, 2

(2006), 152–177.

[33] Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems*. 253–262.

[34] Melanie A Greenberg and Arthur A Stone. 1992. Emotional disclosure about traumas and its relation to health: effects of previous disclosure and trauma severity. *Journal of personality and social psychology* 63, 1 (1992), 75.

[35] Melanie A Greenberg, Camille B Wortman, and Arthur A Stone. 1996. Emotional expression and physical health: Revising traumatic memories or fostering self-regulation? *Journal of personality and social psychology* 71, 3 (1996), 588.

[36] Larry V Hedges and Jack L Vevea. 1998. Fixed-and random-effects models in meta-analysis. *Psychological methods* 3, 4 (1998), 486.

[37] Janise A Hinson and Jane L Swanson. 1993. Willingness to seek help as a function of self-disclosure and problem severity. *Journal of Counseling & Development* 71, 4 (1993), 465–470.

[38] Kasper Hornbæk, Søren S Sander, Javier Andrés Bargas-Avila, and Jakob Grue Simonsen. 2014. Is once enough? On the extent and content of replications in human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3523–3532.

[39] Hsiu-Fang Hsieh and Sarah E Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative health research* 15, 9 (2005), 1277–1288.

[40] George H John and Pat Langley. 2013. Estimating continuous distributions in Bayesian classifiers. *arXiv preprint arXiv:1302.4964* (2013).

[41] Christina J Jones, Mark Hayward, Abigail Brown, Elizabeth Clark, Daniel Bird, Gina Harwood, Crystyn Scott, Angela Hillemann, and Helen E Smith. 2016. Feasibility and participant experiences of a written emotional disclosure intervention for parental caregivers of people with psychosis. *Stress and Health* 32, 5 (2016), 485–493.

[42] Jeffrey H Kahn, Renee M Tobin, Audra E Massey, and Jennifer A Anderson. 2007. Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American journal of psychology* (2007), 263–286.

[43] David M Keating and Stephen A Rains. 2015. Health blogging and social support: a 3-year panel study. *Journal of health communication* 20, 12 (2015), 1449–1457.

[44] Youngsuk Kim. 2005. *Effects of expressive writing among Mexican and Korean bilinguals on social, physical, and mental well-being*. The University of Texas at Austin.

[45] Funda Kivran-Swaine, Jeremy Ting, Jed Richards Brubaker, Rannie Teodoro, and Mor Naaman. 2014. Understanding loneliness in social awareness streams: Expressions and responses. In *Eighth international AAAI conference on weblogs and social media*.

[46] Kitty Klein and Adriel Boals. 2001. Expressive writing can increase working memory capacity. *Journal of experimental psychology: General* 130, 3 (2001), 520.

[47] Heidi E Koschwanez, Ngaire Kerse, Margot Darragh, Paul Jarrett, Roger J Booth, and Elizabeth Broadbent. 2013. Expressive writing and wound healing in older adults: a randomized controlled trial. *Psychosomatic Medicine* 75, 6 (2013), 581–590.

[48] Wouter Marco Kouw and Marco Loog. 2019. A review of domain adaptation without target labels. *IEEE transactions on pattern analysis and machine intelligence* (2019).

[49] Sang Won Lee, Inyeop Kim, Jaehyun Yoo, Sungkyu Park, Bumseok Jeong, and Meeyoung Cha. 2016. Insights from an expressive writing intervention on Facebook to help alleviate depressive symptoms. *Computers in Human Behavior* 62 (2016), 613–619.

[50] Young Ho Lee. 2019. Strengths and Limitations of Meta-Analysis. *The Korean Journal of Medicine* 94, 5 (2019), 391–395.

[51] Stephen J Lepore, Melanie A Greenberg, Michelle Bruno, and Joshua M Smyth. 2002. Expressive writing and health: Self-regulation of emotion-related experience, physiology, and behavior. (2002).

[52] Morton Lieberman. 2007. The role of insightful disclosure in outcomes for women in peer-directed breast cancer groups: a replication study. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer* 16, 10 (2007), 961–964.

[53] Morton A Lieberman and Benjamin A Goldstein. 2006. Not all negative emotions are equal: The role of emotional expression in online support groups for women with breast cancer. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer* 15, 2 (2006), 160–168.

[54] Mingming Liu, Tianli Liu, Xinyu Wang, Nan Zhao, Jia Xue, and Tingshao Zhu. 2019. A linguistic study of Chinese Weibo users who lost their only child. *Death studies* (2019), 1–12.

[55] Carissa A Low, Annette L Stanton, and Sharon Danoff-Burg. 2006. Expressive disclosure and benefit finding among breast cancer patients: Mechanisms for positive health effects. *Health Psychology* 25, 2 (2006), 181.

[56] Qian Lu, Dianhan Zheng, Lucy Young, Marjorie Kagawa-Singer, and Alice Loh. 2012. A pilot study of expressive writing intervention among Chinese-speaking breast cancer survivors. *Health Psychology* 31, 5 (2012), 548.

[57] Luciano L'Abate. 2007. Low-cost approaches to promote physical and mental health. In *Low-cost approaches to promote physical and mental health*. Springer, 3–46.
[58] David Manier and Adelaida Olivares. 2005. Who Benefits from Expressive Writing? Moderator Variables Affecting Outcomes of Emotional Disclosure Interventions. *Counseling & Clinical Psychology Journal* 2, 1 (2005).
[59] Andrew McCallum, Kamal Nigam, et al. 1998. A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization*, Vol. 752. Citeseer, 41–48.
[60] Katelyn YA McKenna and John A Bargh. 2000. Plan 9 from cyberspace: The implications of the Internet for personality and social psychology. *Personality and social psychology review* 4, 1 (2000), 57–75.
[61] Erin L Merz, Rina S Fox, and Vanessa L Malcarne. 2014. Expressive writing interventions in cancer patients: a systematic review. *Health psychology review* 8, 3 (2014), 339–361.
[62] Mohamed Morchid, Richard Dufour, Pierre-Michel Bousquet, Mohamed Bouallegue, Georges Linares, and Renato De Mori. 2014. Improving dialogue classification using a topic space representation and a gaussian classifier based on the decision rule. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 126–130.
[63] Arvind Narayanan and Vitaly Shmatikov. 2010. Myths and fallacies of" personally identifiable information". *Commun. ACM* 53, 6 (2010), 24–26.
[64] Deborah Nazarian and Joshua M Smyth. 2013. An experimental test of instructional manipulations in expressive writing interventions: Examining processes of change. *Journal of Social and Clinical Psychology* 32, 1 (2013), 71–96.
[65] Rebecca J North, Rachel L Meyerson, Danielle N Brown, and Charles J Holahan. 2013. The language of psychological change: Decoding an expressive writing paradigm. *Journal of Language and Social Psychology* 32, 2 (2013), 142–161.
[66] John E Pachankis and Marvin R Goldfried. 2010. Expressive writing for gay-related stress: Psychosocial benefits and mechanisms underlying improvement. *Journal of consulting and clinical psychology* 78, 1 (2010), 98.
[67] Daeun Park, Gerardo Ramirez, and Sian L Beilock. 2014. The role of expressive writing in math anxiety. *Journal of Experimental Psychology: Applied* 20, 2 (2014), 103.
[68] James W Pennebaker. 1997. Writing about emotional experiences as a therapeutic process. *Psychological science* 8, 3 (1997), 162–166.
[69] James W Pennebaker and Sandra K Beall. 1986. Confronting a traumatic event: toward an understanding of inhibition and disease. *Journal of abnormal psychology* 95, 3 (1986), 274.
[70] James W Pennebaker and Cindy K Chung. 2007. Expressive writing, emotional upheavals, and health. *Handbook of health psychology* (2007), 263–284.
[71] James W Pennebaker and Martha E Francis. 1996. Cognitive, emotional, and language processes in disclosure. *Cognition & Emotion* 10, 6 (1996), 601–626.
[72] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71, 2001 (2001), 2001.
[73] James W Pennebaker, Tracy J Mayne, and Martha E Francis. 1997. Linguistic predictors of adaptive bereavement. *Journal of personality and social psychology* 72, 4 (1997), 863.
[74] James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology* 54, 1 (2003), 547–577.
[75] Stephen A Rains. 2014. The implications of stigma and anonymity for self-disclosure in health blogs. *Health communication* 29, 1 (2014), 23–31.
[76] Stephen A Rains and David M Keating. 2011. The social dimension of blogging about health: Health blogging, social support, and well-being. *Communication Monographs* 78, 4 (2011), 511–534.
[77] Stephen A Rains and David M Keating. 2015. Health blogging: An examination of the outcomes associated with making public, written disclosures about health. *Communication Research* 42, 1 (2015), 107–133.
[78] Jane M Richards, Wanda E Beal, Janel D Seagal, and James W Pennebaker. 2000. Effects of disclosure of traumatic events on illness behavior among psychiatric prison inmates. *Journal of Abnormal psychology* 109, 1 (2000), 156.
[79] Josephine P Riddle, HE Smith, and CJ Jones. 2016. Does written emotional disclosure improve the psychological and physical health of caregivers? A systematic review and meta-analysis. *Behaviour research and therapy* 80 (2016), 23–32.
[80] Inna D Rivkin, Julie Gustafson, Ilene Weingarten, and Dorothy Chin. 2006. The effects of expressive writing on adjustment to HIV. *AIDS and Behavior* 10, 1 (2006), 13–26.
[81] Harriet J Rosenberg, Stanley D Rosenberg, Marc S Ernstoff, George L Wolford, Robert J Amdur, Mary R Elshamy, Susan M Bauer-Wu, Tim A Ahles, and James W Pennebaker. 2002. Expressive disclosure and health outcomes in a prostate cancer population. *The International Journal of Psychiatry in Medicine* 32, 1 (2002), 37–53.
[82] Sabirat Rubya and Svetlana Yarosh. 2017. Interpretations of online anonymity in Alcoholics Anonymous and Narcotics Anonymous. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–22.

[83] Koustuv Saha, Sang Chan Kim, Manikanta D Reddy, Albert J Carter, Eva Sharma, Oliver L Haimson, and Munmun De Choudhury. 2019. The language of LGBTQ+ minority stress experiences on social media. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–22.

[84] Sarita Yardi Schoenebeck. 2013. The Secret Life of Online Moms: Anonymity and Disinhibition on YouBeMom. com.. In *ICWSM*. Citeseer.

[85] Judy Hanwen Shen and Frank Rudzicz. 2017. Detecting anxiety through reddit. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality*. 58–65.

[86] Natasha Slesnick and Jillian L Prestopnik. 2004. Office versus home-based family therapy for runaway, alcohol abusing adolescents: Examination of factors associated with treatment attendance. *Alcoholism Treatment Quarterly* 22, 2 (2004), 3–19.

[87] Denise M Sloan, Brian P Marx, and Eva M Epstein. 2005. Further examination of the exposure model underlying the efficacy of written emotional disclosure. *Journal of Consulting and Clinical Psychology* 73, 3 (2005), 549.

[88] Helen E Smith, Christina J Jones, Matthew Hankins, Andy Field, Alice Theadom, Richard Bowskill, Rob Horne, and Anthony J Frew. 2015. The effects of expressive writing on lung function, quality of life, medication use, and symptoms in adults with asthma: a randomized controlled trial. *Psychosomatic medicine* 77, 4 (2015), 429–437.

[89] Joshua M Smyth. 1998. Written emotional expression: effect sizes, outcome types, and moderating variables. *Journal of consulting and clinical psychology* 66, 1 (1998), 174.

[90] Joshua M Smyth and James W Pennebaker. 2008. Exploring the boundary conditions of expressive writing: In search of the right recipe. *British journal of health psychology* 13, 1 (2008), 1–7.

[91] Stefanie P Spera, Eric D Buhrfeind, and James W Pennebaker. 1994. Expressive writing and coping with job loss. *Academy of management journal* 37, 3 (1994), 722–733.

[92] Hannah Stockton, Stephen Joseph, and Nigel Hunt. 2014. Expressive writing and posttraumatic growth: An Internet-based study. *Traumatology: An International Journal* 20, 2 (2014), 75.

[93] Dianna L Stone and Patrick J Rosopa. 2017. The advantages and limitations of using meta-analysis in human resource management research.

[94] S Shyam Sundar, Heidi Hatfield Edwards, Yifeng Hu, and Carmen Stavrositu. 2007. Blogging for better health: Putting the "public" back in public health. *Blogging, citizenship, and the future of media* (2007), 83–102.

[95] Franca Tani, Carole Peterson, and Martina Smorti. 2016. The words of violence: Autobiographical narratives of abused women. *Journal of family violence* 31, 7 (2016), 885–896.

[96] Steven M Toepfer and Kathleen Walker. 2009. Letters of Gratitude: Improving Well-Being through Expressive Writing. *Journal of Writing Research* 1, 3 (2009).

[97] Gabriele Travagin, Davide Margola, and Tracey A Revenson. 2015. How effective are expressive writing interventions for adolescents? A meta-analytic review. *Clinical psychology review* 36 (2015), 42–55.

[98] Sonja Utz. 2015. The function of self-disclosure on social network sites: Not only intimate, but also positive and entertaining self-disclosures increase the feeling of connection. *Computers in Human Behavior* 45 (2015), 1–10.

[99] Maja Van Der Velden and Khaled El Emam. 2013. "Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. *Journal of the American Medical Informatics Association* 20, 1 (2013), 16–24.

[100] Esteban Walker, Adrian V Hernandez, and Michael W Kattan. 2008. Meta-analysis: Its strengths and limitations. *Cleveland Clinic journal of medicine* 75, 6 (2008), 431.

[101] Tao Wang, Markus Brede, Antonella Ianni, and Emmanouil Mentzakis. 2017. Detecting and characterizing eating-disorder communities on social media. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 91–100.

[102] Xinyu Wang, Chunhong Zhang, Yang Ji, Li Sun, Leijia Wu, and Zhana Bao. 2013. A depression detection model based on sentiment analysis in micro-blog social network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 201–213.

[103] Markus Wolf, Jan Sedway, Cynthia M Bulik, and Hans Kordy. 2007. Linguistic analyses of natural written language: Unobtrusive assessment of cognitive style in eating disorders. *International journal of eating disorders* 40, 8 (2007), 711–717.

[104] Jeannie Wright and Man Cheung Chung. 2001. Mastery or mystery? Therapeutic writing: A review of the literature. *British Journal of Guidance & Counselling* 29, 3 (2001), 277–291.

[105] Diyi Yang, Zheng Yao, and Robert Kraut. 2017. Self-disclosure and channel difference in online health support groups. In *Proceedings of the... International AAAI Conference on Weblogs and Social Media. International AAAI Conference on Weblogs and Social Media*, Vol. 2017. NIH Public Access, 704.

[106] Diyi Yang, Zheng Yao, Joseph Seering, and Robert Kraut. 2019. The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.

[107] Svetlana Yarosh. 2013. Shifting dynamics or breaking sacred traditions? The role of technology in twelve-step fellowships. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3413–3422.

A LIST OF PAPERS IN THE FINAL POOL

Title	Author	Year	Participants
Cognitive, emotional, and language processes in disclosure	Pennebaker	1996	student
Effects of disclosure of traumatic events on illness behavior among psychiatric prison inmates	Richards	2000	prisoner
Expressive writing can increase working memory capacity	Klein	2001	student
Praying about difficult experiences as self-disclosure to God	VandeCreek	2002	student
Transformational processing: Healthy identity functioning in written narratives of emotionally challenging life events	Smith	2003	student
Brief Report: Psychological Impact of Writing About Abuse or Positive Experiences	Antal	2005	student
Getting to the heart of the matter: Written disclosure, gender, and heart rate	Epstein	2005	student
Language use in imagined dialogue and narrative disclosures of trauma	Burke	2006	student
Cognitive-behavioral models of emotional writing: A validation study	Guastella	2006	student
Expressive disclosure and benefit finding among breast cancer patients: Mechanisms for positive health effects	Low	2006	patient
The Effects of Expressive Writing on Adjustment to HIV	Rivkin	2006	patient
Measuring Engagement as a Moderator within an Expressive Writing Intervention for Smokers	Stone	2006	smoker
Affectionate writing reduces total cholesterol: Two randomized, controlled trials	Floyd	2007	student
Does Expressive Writing Reduce Stress and Improve Health for Family Caregivers of Older Adults?	Mackenzie	2007	caregiver
Telling and the remembered self: Linguistic differences in memories for previously disclosed and previously undisclosed events	Pasupathi	2007	student
Does altering the writing instructions influence outcome associated with written disclosure?	Sloan	2007	student
Changes in mood predict disease activity and quality of life in patients with psoriasis following emotional disclosure	Vedhara	2007	patient
Linguistic analyses of natural written language: Unobtrusive assessment of cognitive style in eating disorders	Wolf	2007	patient
Variations in the spacing of expressive writing sessions	Chung	2008	student
Are alexithymia and emotional characteristics of disclosure associated with blood pressure reactivity and psychological distress following written emotional disclosure?	Oconnor	2008	adult

Healing the wounds of organizational injustice: Examining the benefits of expressive writing	Barclay	2009	student
The Effects of Pennebaker's Writing Paradigm on Physical and Emotional Distress: An Exploration of Narrative Content and Moderating Variables	Kearns	2009	student
Assessing coping strategies by analysing expressive writing samples	Lee	2009	student
Expressive writing with university students with disabilities	Lotze	2009	disabled
Potential benefits of expressive writing for male college students with varying degrees of restrictive emotionality.	Wong	2009	student
The effects of web-based interactive emotional disclosure on stress and health: A randomized, controlled study	Beyer	2010	PTSD
Therapeutic writing as an intervention for symptoms of bulimia nervosa: effects and mechanism of change	Johnston	2010	student
Assessing coping strategies by analysing expressive writing samples	Lee	2010	student
Participant experiences of a written emotional disclosure intervention in asthma	Theadom	2010	patient
Coping with stressful events: Use of cognitive words in stressful narratives and the meaning-making process	Boals	2011	student
RU stressed?: the physical and psychological effects of odor exposure and writing about stressful experiences	Castellanos	2011	student
Does emotional disclosure about stress improve health in rheumatoid arthritis? Randomized, controlled trials of written and spoken disclosure	Lumley	2011	patient
Linguistically-tailored video feedback increases total and positive emotional expression in a structured writing task	Owen	2011	adult
An investigation of the efficacy of online expressive writing for trauma-related psychological distress in Hispanic individuals	Hiral	2012	student
Alexithymia, emotional intelligence, and their relation to word usage in expressive writing	Pluth	2012	patient
Alexithymia, emotional intelligence, and their relation to word usage in expressive writing	Hanson	2013	adult
An Experimental Test of Instructional Manipulations in Expressive Writing Interventions: Examining Processes of Change	Nazarian	2013	student
The Language of Psychological Change: Decoding an Expressive Writing Paradigm	North	2013	student
Computer-based written emotional disclosure: the effects of advance or real-time guidance and moderation by Big 5 personality traits	Beyer	2014	student
Expressive writing and posttraumatic growth: An Internet-based study	Stockton	2014	patient

Pilot trial of an expressive writing intervention with HIV-positive methamphetamine-using men who have sex with men	Carrico	2015	patient
Randomised controlled trial of expressive writing and quality of life in men and women treated for colon or rectal cancer	Lepore	2015	patient
The Effects of Expressive Writing on Lung Function, Quality of Life, Medication Use, and Symptoms in Adults With Asthma: A Randomized Controlled Trial	Smith	2015	patient
Feasibility and participant experiences of a written emotional disclosure intervention for parental caregivers of people with psychosis	Jones	2016	caregiver
The effects of expressive writing before or after punch biopsy on wound healing	Robinson	2017	patient
The Roles of Self-Affirmation and Emotional Disclosure in Promoting Adjustment to Chronic Financial Stress: An Experimental Investigation	Bauer	2018	student
The effect of expressive writing on the error-related negativity among individuals with chronic worry	Schroder	2018	student

Table 5. List of 47 papers in the final pool.

B LIST OF PAPERS INCLUDED IN META-ANALYSIS OF POSITVE EMOTION WORDS

Author	Year	Expressive Writing			Non-expressive Writing		
		mean (%)	*sd (%)*	*N*	*mean (%)*	*sd (%)*	*N*
Richards	2000	2.36	1.03	36	2.80	1.37	29
Klein	2001	3.00	0.65	35	1.30	0.59	36
Antal	2005	2.27	0.91	16	1.19	1.27	26
Epstein	2005	0.90	0.56	51	0.48	0.23	43
Burke	2006	2.10	0.80	57	1.00	0.70	63
Low	2006	2.94	1.12	20	1.41	1.12	16
Rivkin	2006	2.67	1.21	22	1.90	1.15	23
Floyd	2007	4.21	1.89	15	2.71	1.40	15
Mackenzie	2007	2.66	1.00	14	1.19	0.73	13
Sloan	2007	2.08	0.63	28	1.12	0.44	27
Vedhara	2007	1.90	1.10	31	1.43	0.96	28
Oconnor	2008	2.25	0.88	43	0.75	0.42	44
Kearns	2009	3.40	0.74	38	1.54	0.49	33
Lotze	2009	3.17	0.97	24	1.63	0.90	26
Wong	2009	4.00	1.06	79	1.95	0.68	74
Beyer	2010	1.90	0.54	40	1.84	0.69	36
Theadom	2010	2.46	1.50	18	1.18	1.27	18
Boals	2011	2.51	1.17	89	2.12	1.30	88
Castellanos	2011	2.81	1.22	40	1.28	0.92	40
Lumley	2011	2.54	0.92	43	2.09	0.94	21
Hiral	2012	2.45	1.37	54	2.16	1.60	50
Hanson	2013	2.92	1.86	8	1.65	0.81	7
Nazarian	2013	2.72	0.99	34	2.31	1.02	33
Beyer	2014	1.90	0.54	41	1.85	0.68	40
Stockton	2014	2.21	1.14	14	1.73	1.11	10
Carrico	2015	3.88	0.87	12	3.67	3.32	11
Lepore	2015	2.40	0.92	101	2.68	0.78	92
Smith	2015	2.23	1.02	55	2.02	1.60	65
Jones	2016	1.93	1.08	12	1.78	0.73	12
Robinson	2017	3.09	1.06	31	1.82	0.87	33
Bauer	2018	2.85	0.73	36	1.26	0.52	37
Schroder	2018	3.75	1.67	21	1.47	0.98	19

Table 6. List of 32 papers in the meta-analysis of positive emotion words. % denotes percentage of total number of words. *N* denotes number of participants' essays in the writing condition. In total, there are 1158 essays included in the meta-analysis of positive emotion words for the expressive writing condition and 1108 essays for the control condition.

C LIST OF PAPERS INCLUDED IN META-ANALYSIS OF NEGATIVE EMOTION WORDS

Author	Year	Expressive Writing			Non-expressive Writing		
		mean (%)	*sd (%)*	*N*	*mean (%)*	*sd (%)*	*N*
Richards	2000	2.80	1.24	36	1.39	1.15	29
Klein	2001	2.00	0.71	35	0.34	0.65	36
Antal	2005	2.97	1.30	16	0.51	0.46	26
Epstein	2005	1.24	0.48	51	0.29	0.25	43
Burke	2006	2.10	0.90	57	0.50	0.30	63
Low	2006	2.24	0.58	20	0.98	0.60	16
Rivkin	2006	2.05	0.96	22	0.89	0.92	23
Floyd	2007	1.01	0.55	15	0.62	0.47	15
Mackenzie	2007	2.16	0.90	14	0.37	0.40	13
Sloan	2007	3.05	1.12	28	0.56	0.33	27
Vedhara	2007	2.23	1.16	31	0.45	0.45	28
Oconnor	2008	2.98	0.99	43	0.32	0.27	44
Barclay	2009	3.07	1.45	25	0.52	1.45	25
Kearns	2009	3.02	0.79	38	0.55	0.27	33
Lotze	2009	2.53	1.20	24	0.60	0.73	26
Wong	2009	1.88	0.70	79	1.08	0.46	74
Beyer	2010	2.62	0.70	40	0.87	0.62	36
Theadom	2010	2.25	1.86	18	0.29	0.62	18
Boals	2011	3.60	1.54	89	2.87	1.45	88
Castellanos	2011	3.02	1.45	40	0.42	0.49	40
Lumley	2011	3.16	1.23	43	0.66	0.88	21
Hiral	2012	4.10	1.56	54	2.54	1.46	50
Hanson	2013	4.27	1.73	8	0.53	0.35	7
Nazarian	2013	3.53	1.18	34	0.57	0.40	33
Stockton	2014	4.00	1.56	14	1.05	0.79	10
Carrico	2015	2.35	0.67	12	0.78	0.66	11
Lepore	2015	2.16	0.81	101	0.64	0.47	92
Smith	2015	2.82	1.26	55	0.46	0.44	65
Jones	2016	2.91	1.14	12	0.72	0.46	12
Robinson	2017	3.55	1.92	31	0.50	0.40	33
Bauer	2018	2.77	0.74	36	0.66	0.33	37
Schroder	2018	2.52	1.19	21	0.52	0.60	19

Table 7. List of 32 papers in the meta-analysis of negative emotion words. % denotes percentage of total number of words. *N* denotes number of participants' essays in the writing condition. In total, there are 1142 essays included in the meta-analysis of negative emotion words for the expressive writing condition and 1093 essays for the control condition.

D LIST OF PAPERS INCLUDED IN META-ANALYSIS OF CAUSAL WORDS

Author	Year	Expressive Writing			Non-expressive Writing		
		mean (%)	*sd (%)*	*N*	*mean (%)*	*sd (%)*	*N*
Klein	2001	1.10	0.35	35	0.53	0.35	36
Burke	2006	1.00	0.50	57	0.60	0.40	63
Rivkin	2006	1.41	0.81	22	0.60	0.60	23
Barclay	2009	2.30	0.80	25	1.13	0.80	25
Kearns	2009	2.15	0.44	38	1.04	0.29	33
Lotze	2009	2.00	0.94	24	0.97	0.81	26
Wong	2009	1.30	0.45	79	1.57	0.67	74
Theadom	2010	0.87	0.88	18	0.18	0.31	18
Castellanos	2011	2.36	1.07	40	0.98	0.70	40
Lumley	2011	1.45	0.50	43	0.95	0.55	21
Stockton	2014	1.72	0.92	14	1.12	0.64	10
Smith	2015	1.03	0.67	55	0.27	0.34	65
Jones	2016	0.79	0.41	12	0.35	0.29	12
Bauer	2018	2.48	0.66	36	1.36	0.56	37
Schroder	2018	2.89	1.27	21	0.95	0.65	19

Table 8. List of 15 papers in the meta-analysis of causal words. *%* denotes percentage of total number of words. *N* denotes number of participants' essays in the writing condition. In total, there are 519 essays included in the meta-analysis of causal words for the expressive writing condition and 502 essays for the control condition.

E LIST OF PAPERS INCLUDED IN META-ANALYSIS OF INSIGHT WORDS

Author	Year	Expressive Writing			Non-expressive Writing		
		mean (%)	*sd (%)*	*N*	*mean (%)*	*sd (%)*	*N*
Klein	2001	2.40	0.65	35	0.74	0.41	36
Antal	2005	3.07	1.10	16	0.92	0.62	26
Burke	2006	2.00	0.60	57	1.10	0.60	63
Rivkin	2006	2.78	1.13	22	1.27	0.83	23
Kearns	2009	3.65	0.61	38	0.74	0.34	33
Lotze	2009	3.97	1.48	24	1.20	1.08	26
Wong	2009	3.94	1.14	79	2.92	1.11	74
Theadom	2010	2.35	1.44	18	0.45	0.85	18
Castellanos	2011	3.15	1.28	40	0.40	0.56	40
Lumley	2011	2.63	0.89	43	0.91	1.30	21
Beyer	2014	2.61	0.71	41	1.23	0.59	40
Stockton	2014	3.58	1.53	14	1.45	1.33	10
Smith	2015	2.60	1.13	55	0.56	0.49	65
Jones	2016	2.13	0.71	12	1.26	0.75	12
Robinson	2017	3.37	1.06	31	0.88	0.54	33
Bauer	2018	3.00	0.74	36	1.05	0.41	37
Schroder	2018	4.30	1.93	21	0.99	1.21	19

Table 9. List of 17 papers in the meta-analysis of insight words. % denotes percentage of total number of words. *N* denotes number of participants' essays in the writing condition. In total, there are 582 essays included in the meta-analysis of insight words for the expressive writing condition and 576 essays for the control condition.

F CODING INSTRUCTIONS

*Please read each sentence closely and use your best judgment to decide if a sentence is a key sentence that signals expressive writing. A key sentence must be self-disclosure of thoughts and feelings about health journeys or other life struggles. Specifically, a key sentence writes about: 1) one's general feelings, moods, or mental states that are not directed at, related to, or preceded by a specific event or object; 2) one's thoughts and emotions about stressful events; 3) one's reflection on the health journey (e.g. what he/she has learned, what he/she would like to change); or 4) one's coping with struggles and difficulties (e.g. discovery of meaning, accepting life, taking a self-distanced perspective, believing and praising God). A sentence is **NOT** a key sentence, if it writes about: 1) facts, descriptions, plans, or schedule; 2) one's thoughts and emotions about mundane activities, social events, or everyday objects (e.g. we visited some old friends and felt happy); 3) others' thoughts and emotions (e.g. I know she must be upset); 4) content quoted from books, bibles, or other sources; or 5) common expressions in online health communities (e.g. requesting prayers, messages, visits, or other support, expressing gratitude, sending good wishes). Mark each key sentence for future reference. Label an update as expressive writing (denote as 1), if it contains at least one key sentence. Label an update as non-expressive writing (denote as 0), if it does not have any key sentence.*

Received June 2020; revised October 2020; accepted December 2020