Unraveling Abstinence and Relapse: Smoking Cessation Reflected in Social Media

Elizabeth L. Murnane[1,2] **Scott Counts**[1]
elm236@cornell.edu counts@microsoft.com
[1]Microsoft Research, [2]Cornell University Information Science

ABSTRACT
Analysis of smokers' posts and behaviors on Twitter reveals factors impacting abstinence and relapse during cessation attempts. Combining automatic and crowdsourced techniques, we detect users trying to quit smoking and analyze tweet and network data from a sample of 653 individuals over a two-year window of quitting. Guided by theory and practice, we derive behavioral, social, and emotional measures to compare users who abstain and relapse. We also examine the cessation process, demonstrating that Twitter can help chronicle how some people go about quitting. Among other results, we show that those who fail in their smoking cessation are far heavier posters and use relatively less positive language, while those who succeed are more social in both network ties and in directed communication. We conclude with insights on how intelligent intervention systems can harness these signals to provide tailored behavior change support.

Author Keywords
Smoking; cessation; health; behavior; social media; Twitter

ACM Classification Keywords
H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces

INTRODUCTION

Tobacco smoking is the leading cause of preventable death and the leading form of chemical dependence in the U.S, resulting in over 400,000 deaths annually [12]. According to the latest report [11] by the Centers for Disease Control and Prevention (CDC), there are nearly 44 million smokers in the U.S. alone. Importantly and encouragingly, over 68% of smokers report a desire to quit and over 50% have attempted to do so for at least one day. However, relapse is common, and only a minority of smokers are able to permanently maintain abstinence [8].

There is considerable recent evidence of the efficacy of various intervention strategies, including individual or group therapy, physician care, and self-help interventions [16, 26]. However, the former programs are not accessible, affordable, or utilized beyond a minority of smokers; and self-help strategies must be closely tailored to individual habits, circumstances, and progress to be effective [41].

Digital cessation tools, such as phone-based monitoring and assistance applications, hold tremendous potential to overcome these challenges to widespread adoption by empowering a range of users with information, strategies, and self-awareness when attempting to quit. While such tools are appearing, they have major shortcomings as well. Studies of these tools find low adherence to established clinical practice guidelines for treating tobacco use and dependence [1]. This same prior work and our own evaluation also find that these tools' support is not personalized, is not based on individual motivations, and does not adapt during the cessation process, for instance to detect or pre-empt user struggles and setbacks. Rather, their approaches are typically one size-fits all and use game-ified elements to maintain engagement instead of the customized support necessary to achieve successful cessation outcomes.

In this work, we identify the posting and network data of smokers on Twitter as a way to observe and assess key personal and social traits and behaviors relevant to providing this sort of tailored support. We aim to leverage this increasingly abundant data to develop methods for assessing and predicting a smoker's likelihood of remaining abstinent during the cessation process. Our underlying motivation is creating a new class of personalized and sustainable intervention tools for health-related behavior change that engage with a broader spectrum of users.

BACKGROUND

Goal Setting, Behavior Change, and Smoking Cessation

Seminal theory, recent refinements, and experimental research establish that a combination of personal, behavioral, and environmental factors influence why an individual sets a goal, performs positively or negatively during its pursuit, and ultimately reaches success or failure [6, 7, 28]. The Transtheoretical Model (TTM) [35] provides us a conceptual framework with which to evaluate an individual's readiness to embark on such a goal and monitor progress through stages of behavior change. More specifically, individuals do not have any near-term intention to quit when in *Pre-Contemplation*; begin to more fully realize the cons of the unhealthy behavior in *Contemplation*; continue evaluating the positive impacts changing could have on oneself and others as they progress

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada

http://dx.doi.org/10.1145/2556288.2557145

through *Preparation*; and rely on commitments, conditioning, environmental controls, and social support as they reach *Action* (i.e., modify behavior) and sustain long term *Maintenance* [34, 35]. One contribution of our current work is developing social media-based measures that capture such aspects of TTM, which can then be embedded into tools for many types of behavior change.

Numerous studies over the years have tested these models with smoking cessation interventions. Findings generally support the model and confirm that stage of change variables positively predict abstinence [15, 36], including for adolescents [38] and for smokers with varying mindsets about quitting [41]. However, some studies have met criticisms due to non-representative recruitment or lack of longitudinal follow-up beyond 1 to 24 months [4, 41].

Text and Social Network Analysis

Advances in psycholinguistics have shown the effectiveness of using the text people write to evaluate emotional, psychological, cognitive, and behavioral attributes relevant to the aforementioned theoretical constructs. Sentiment analysis techniques and tools such as the Linguistic Inquiry Word Count (LIWC) [32] have been well validated. Through such text analysis, researchers have found strong correlations between language use and physical health, cognitive processes, and mental state [10, 31, 39]. Linguistic analysis has also been used to model emotions and affective states from text [9] such as Twitter posts [25].

Other recent work analyzes the size and structure of social networks and finds that the connections and interactions of a user can strongly influence her own health-related behaviors and goals [5, 17]. In this setting, a person is exposed to temptations as well as others' achievements, which can have negative or positive impacts on her own motivation, expectations, and effort [2, 6]. Further, the social context creates chances to seek and receive external support, with encouragement or constructive feedback improving performance [27] and negative messaging having even stronger effects in the opposite direction [6].

Twitter and Personal Health

Data from Twitter in particular are attractive for connecting users' online generated content and social relationships with personal health. Tweets are naturally expressed in a user's own voice, permitting the kinds of linguistic analysis previously mentioned. Additionally, Twitter's social setting offers opportunity to study the accountability, exposure to success, and peer support important to goal formation, motivation, and achievement as just discussed. Finally, tweets are posted publicly, spontaneously, and over extended periods of time. This enables low-cost, large-scale, and longitudinal collection of data that exposes more realistic patterns of activity than traditional survey methods, which face challenges of low response rates and difficulty ensuring veracity of respondents' reported behavior.

With respect to the health domain, analysis of Twitter data has proved a promising way to diagnose and better understand health related behaviors and disorders such as insomnia [24], mental health [13], and fitness [40].

Regarding smoking specifically, research has very recently recognized Twitter as an ideal medium through which to promote smoking cessation programs [33]. However, the analysis of posting behavior has gone untapped as a way to identify smokers, better understand the smoking and abstinence processes, and determine signals of an attempt to quit or a need for cessation support.

METHOD

Identifying Smokers

Our first step was to identify Twitter users who are smokers with an intention to quit. Our qualitative examination of cessation forums, self-help materials, and tweets revealed key terms and topics relevant to smoking cessation. We used these to build queries (e.g., *+smok*|+cig* +[quit* | smoke free | cessation] [today] –ago [official*])* for the Twitter Firehose (made available to us via an agreement with Twitter) to retrieve *cessation-event* tweets posted between June 1, 2011 and June 1, 2012. This time window allowed us to collect a year's worth of a user's tweets both before and after an announcement to quit, letting us look for pre-cessation motives as well as for changes in behavior, emotion, or social dynamics following the cessation-event.

From these cessation-event tweets, we eliminated retweets (since the original message is posted by a different user), non-English tweets or those by users whose profiles do not specify language as English, tweets by users with under 100 lifetime tweets, tweets by known advertising accounts or containing defined advertisement keywords, and tweets referring to smoking-substances other than tobacco. From the resulting 33,420 posts, we sampled 2000 randomly.

We removed any personally identifiable information from the text, for example normalizing mentions to @username, and then used Mechanical Turkers to manually verify each tweet as a genuine announcement of an intention to quit. Each tweet was labeled by 3 workers required to be Masters and live in the U.S. The judgments of any annotator who did not specify a minimum familiarity level with Twitter or who did not take a minimum amount of time to complete the task were disregarded as invalid. To be conservative, we eliminated tweets not unanimously labeled as authentic announcements to quit, resulting in 733 unique users and cessation-event tweets. Fleiss kappa = 0.7613, which falls within a generally accepted range of rater agreement. The following are examples of verified cessation-event tweets:

- *Just smoked my last cigarette ever!!! I officially quit!*
- *Today I am going to quit smoking. Wish me luck*
- *Last cigarette ever today. #promiseonmylife #bestfeeling*

Having identified a set of smokers intending to quit, we collected these users' publicly available profile and network information. We also again iterated the Twitter Firehose to collect their timeline of tweets posted one year prior to their cessation-event and one year after. We identified smoking-relevant tweets using our smoking keywords originally used to retrieve cessation-events. We used Mechanical Turk to verify these tweets as relevant to smoking and the user's own habit, in which case Turkers also coded the Behavior Change Process variables described in the next section. Turkers received labels, definitions, and example tweets for all possible cessation motives, methods, and TTM stages; and like all our MTurk tasks, we disregarded tweets not coded unanimously by 3 Turkers (Fleiss kappa = 0.7615).

Finally, we eliminated any user who referenced non-tobacco smoking-substances more than a threshold amount and any user with zero smoking-relevant tweets after the cessation, leaving a sample of 653 users and 5,309,510 of their tweets (a median of 3,236 tweets per user).

Measures

As discussed earlier, past research suggests a number of determinants that predict whether a person will successfully achieve a goal to quit smoking, and we now describe our attempt to use Twitter data to operationalize those variables plus derive new measures. While we explore a wide array of variables, it is important to note that our intention is not to simply throw any and all possible metrics at the wall but rather to carefully use prior theory and experimentation as a guide in constructing meaningful features.

Response Variable

Outcome represents a user's ultimate *Survival* or *Relapse*. We measure outcome as a clinician would in an offline intervention, recording whether a user is remaining abstinent as of the last available tweet (i.e., the final "assessment") or has relapsed. Recall that we used Mechanical Turk to label the stage of behavior change indicated by a smoking-relevant tweet, and we assign a survival or relapse assessment using this label. We refer to users who do and do not remain abstinent during cessation attempts as **Survivors** and **Relapsers**, respectively, and the following are examples of final assessment tweets:

Survivors	**Relapsers**
• *Congratulations to me, still smoke free :)*	• *Day 26: Broke down and bought a pack of smokes last weekend. Smoked the last one today.*
• *@username nope i don't smoke anymore*	• *Well, tried to quit smokin tobacco but..had a fucked up day*
• *first few weeks were hard but I haven't craved a cig in months*	• *So day 3 of not smoking is about to get cut short..i can't do it lol*

Explanatory Variables

We organize our predictor variables within 4 main categories: online activity, social network structure and interactions, emotional state, and behavior change process. We record measures of these variables according to all of a user's tweets as well as on smoking-relevant tweets only.

Activity variables assess baseline activity level as well as smoking fixation and cessation perseverance:

- Tweet Volume: Total number of tweets a user posts in the 1 year before and 1 year after his/her cessation-event
- Burstiness: Maximum number of tweets a user posts in any single hour
- Frequency: Number of hours between successive pairs of tweets a user posts

Social variables capture a user's support system, incoming and outgoing attention, and passive and active relationships:

- Friends: Number of other users a user follows
- Followers: Number of other users a user is followed by
- Tweets with At-Mentions: Number of tweets containing at least 1 @username mention
- Unique Mentions: Number of unique users @mentioned

Personal and Emotional variables collect demographic, psychological, and affective information about the user:

- Location: The location specified in the user's profile
- Sentiment Intensity Rate: A measure of how intensely positive or negative a user's emotions are, computed using ANEW [9] and LIWC [31] as (per Hutto et al. [23]) the ratio of the sum of the valence intensity of positive or negative language used in tweets to the total number of tweets in a period

Behavior Change Process variables are designed to model a user's cessation process in order to evaluate whether or not she is exhibiting behavioral signals known to correlate with successful outcomes. As described earlier, Mechanical Turk coders provided Motive, Treatment, and TTM stage.

- Cessation Date: Date when a user announces attempt to quit smoking (i.e., posting date of cessation-event tweet)
- Motive to Quit: The motivation triggering a user to quit
- Treatment: Any cessation strategy used to stay abstinent
- PreContemplation-, Contemplation-, Preparation-, and Abstinence-Volume: The number of tweets indicating a user is in the corresponding TTM behavior change stage

RESULTS

We begin with comparisons to CDC statistics both in order to understand the nature of our dataset as well as to buttress our assumptions that we are working with a realistic and representative sample of the U.S. population of smokers. We then contrast our 175 Survivors and 344 Relapsers according to the measures presented in the previous section.

Alignment with CDC Reports

We find that the locations of smokers across the country, the gender of smokers, and abstinence rates closely align with those reported by the CDC.

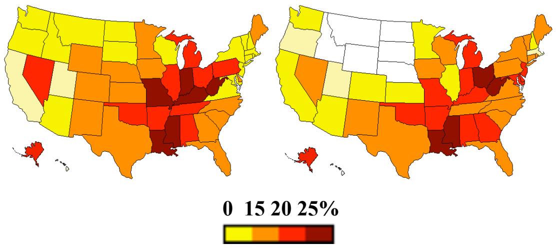

Figure 1. Smoking prevalance according to CDC reports (left) and according to Twitter profiles (right)

Figure 1 illustrates the concentration of smokers in US states according to the CDC [3] and according to the Location field in Twitter profiles of users in our dataset. Darker colors correspond to higher smoking prevalence, and white states are those not represented in our data.

Performing gender classification [13] on users in our dataset, Table 1 shows results very similar to national statistics of tobacco usage by men and women [3].

	CDC	Twitter
Men	54%	59%
Women	46%	41%

Table 1. Proportion of smokers that are men or women according to data from CDC and from Twitter

Finally, over time we see that trends in data on Twitter closely resemble what clinicians encounter when making longitudinal health assessments, as compared in Figure 2. On the left we see the combined abstinence rate from 17 studies of smokers [22], and on the right the same plot of the number of months users in our own dataset remain abstinent (either before relapsing for Relapsers or until the latest assessment for Survivors). An important contrast does exist here though: obtaining longitudinal data via Twitter is much less difficult than in such offline settings, where follow-ups are typically only performed at points between 1 and 12 months and rarely after 24 months [38].

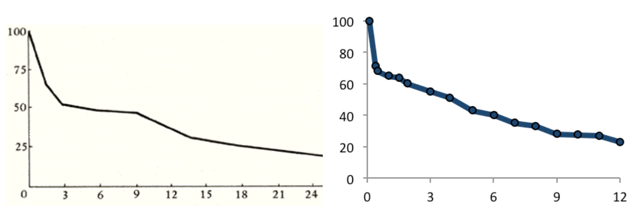

Figure 2. Percentage of people vs. months of abstinence from 17 combined clinical studies (left) and from Twitter dataset (right)

These alignments help verify the soundness of our data, and they also demonstrate that easily-accessible web data could substitute for or enrich data from surveys and other conventional yet burdensome methods of collection.

Activity, Sociality, & Emotion

Next we present our analyses that explore meaningful differences in the online behaviors, social dynamics, and affective states of smokers who successfully remain abstinent (Survivors) and those who relapse and begin smoking again (Relapsers). Our later Discussion section explains how personalized and context-aware tools could capitalize on our findings. We compare these two groups by reporting on the difference in the median values of our measurements computed on the following sets of tweets:

- **All_Before; All_After**: All of a user's tweets posted up to 1 year before or 1 year after his/her cessation-event
- **Smoke_Before; Smoke_After**: Only the tweets of a user that are annotated as smoking-relevant, posted up to 1 year before or 1 year after his/her cessation-event

We removed outliers greater than 3 standard deviations above the mean posting volume, and all comparisons were done on medians using Wilcoxon sign-rank tests.

Activity variables

Striking is a significantly higher activity level for Relapsers compared to Survivors across all variables, as summarized in Table 2. Specifically, Relapsers tweet over 3 times more than Survivors before cessation and nearly 5 times more after cessation, and Relapsers' posting bursts produce twice as many tweets as Survivors', both before and after cessation. Also, posting burstiness and frequency increase after quitting for Relapser yet decrease for Survivors.

	S	**R**
Vol_B* **	412	1243
Vol_A* **	771	3551
Burst_B* **	4.46	10.12
Burst_A* **	4.28	10.94
Freq_B* **	9.91	3.56
Freq_A* **	11.25	2.70

Table 2. Median values of activity measures computed on tweets of (S)urvivors and (R)elapsers, (B)efore and (A)fter cessation. Significant differences in these medians indicated in the first column. (* p < 0.01, ** p < 0.001, *** p < 0.0001)

These significantly higher levels of overall tweeting activity by Relapsers compared to Survivors may be explained by the fact that heavier Twitter use is a manifestation of the impulsiveness and sensation seeking that is characteristic of heavier smokers [33]. Looking at these measures for smoking-relevant tweets only, we see that Relapsers also mention smoking significantly more than Survivors both before and after cessation ($p<0.0001$). This may be a sign that Relapsers have more severe addictions that are difficult to break given that stronger addiction levels correlate with thinking about smoking more frequently during the day [19]. Tweets posted by two of the most active users in our dataset, both Relapsers, illustrate this severity of addiction (e.g., "*I had 8 cigarettes earlier in the space of 30 minutes. Whoops*") and such preoccupation with smoking (e.g., "*I'm tryna get myself to stop smoking but I know it's not gonna happen. I'm already thinking about smoking!*").

Next, looking more closely at when Relapsers and Survivors actually make these posts, we notice temporal differences in posting behavior as well, with Relapsers making more of their smoking mentions at night compared to Survivors during the day, as illustrated in Figure 3. We see at least two potential explanations for this.

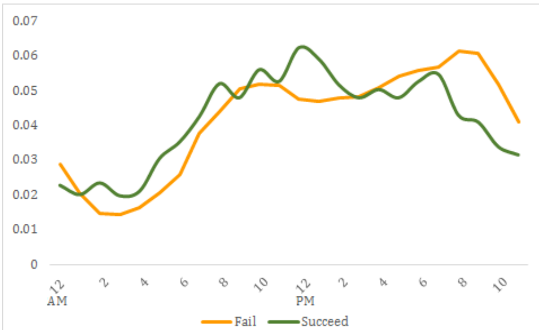

Figure 3. Time of day at which users tweet about smoking

First, research has shown that stress levels often peak at night and that depressed individuals are more active at night on Twitter [13]. We therefore suggest that those users suffering from stress or depression may be more prone to turn to cigarettes for coping or have more trouble resisting the relief from stress smoking provides (e.g., "*im really considering smoking tonight bcause im so stressed*").

Second, research also shows that Relapsers struggle more with temptation [14] and that living with other smokers negatively predicts successful cessation outcomes [18, 29]. We hypothesize that Relapsers encounter situations that threaten abstinence more at night. For example, we see Relapsers being tempted at social outings, especially if the individual is drinking or others nearby are smoking, (e.g., "*outside the club and guy beside me smoking makes me wanna*"), as well as when at home at the end of the day and exposed to other smokers (e.g., *"my mom smokes i stay with her she does not respect me trying to stop :\"*).

Social variables

Next, we examine smokers' social relationships and interactions. In order to avoid Relapsers dominating the raw measures simply due to their massive posting volume, we normalize all social variables by posting volume.

Table 3 shows the differences in friends per tweet and followers per tweet between Survivors and Relapsers, with Survivors benefitting from significantly more friends and followers. These connections enable social outreach and attention, key ingredients in maintaining motivation and performance [27] and navigating the cessation process [8, 29]. Looking to the data, we find many examples of Survivors reaching out, such as "*Starting the patch today. Everyone please support me on the road to quitting smoking*" and "*Ok I started a really big challenge yesterday... I quit smoking! I may need some help from you guys in the upcoming days/weeks*".

	S-R Before	S-R After
Friends/Tweet	0.0948***	0.1340***
Followers/Tweet	0.0400**	0.0608***

Table 3. Differences in the median values of social network structure variables Before and After cessation for (S)urvivors and (R)elapsers. (* p < 0.01, ** p < 0.001, *** p < 0.0001)

Survivors' number of connections also increases from before to after cessation, while Relapsers lose connections. We hypothesize that this is due to Relapsers becoming more socially withdrawn and hostile. For example, consider this Relapser's increasingly anti-social behavior:

- *Day 2 of not smoking #bittersweet*
- *I quit smoking yesterday and everyone is pissing me off!*
- *Day 3 without a cig. Ooo I'm about to shoot someone*

Along the same lines, it also seems the overall emotional tenor of Relapsers' conversations becomes more negative after quitting, perhaps repelling new relationships and even breaking existing ones. Consider the following examples: "*Day 2 no smoking. I hate everyone*" and "*I need a cigarette to calm down... im about to punch something #nojokes*". We discuss these kinds of affective changes further in the following sub-section on emotional variables.

Followers and friends represent more passive subscription and broadcast style relationships, so we also collect @mentions to capture active, personal interactions. Table 4 shows that Survivors have significantly more @mentions per tweet than Relapsers both before and after cessation.

	S-R All_ Before	S-R All_ After	S-R Smoke_ Before	S-R Smoke_ After
% Tweets with @Mentions	0.0529*	0.0396*	-0.0007*	-0.0005*
Unique @Mentions/Tweet	0.0675*	0.0313*	-0.0007*	-0.0007*

Table 4. Differences in the median values of social interaction variables Before and After cessation for (S)urvivors and (R)elapsers in all tweets (All) and in only smoking-relevant tweets (Smoke). (* p < 0.05, ** p < 0.001, *** p < 0.0001)

At the same time, we see that Survivors make significantly fewer @mentions in their smoking-relevant tweets than Relapsers. While Survivors are overall more socially engaged, one possibility is that when tweeting about smoking, they reach out more selectively, consciously targeting closer and more trusted connections for information (e.g., *"@username what is your opinion on vapor cigarettes? Do you think they make it easier to quit?"*) or for emotional encouragement and camaraderie (e.g., "*I quit smoking today! A promise I made to @username & I intend on keeping this promise*").

Emotion variables

Research shows that high positivity and moderate negativity correlate with health [31]. Table 5 shows the differences in the intensity of positive and negative language (measured per Hutto et al. [23]) used by Survivors and Relapsers. We see Survivors speaking more positively before and after quitting and more positively when referring to smoking after quitting. We also see that Relapsers' language is only slightly more emotionally negative than that of Survivors, suggesting that Survivors manage emotional balance without repressing negative feelings, which can result in subsequent health problems [31]. Notice how this Relapser has a largely negative attitude towards cessation: "*So grouchy today, who woulda knew that quitting smoking would be so hard??*" while this Survivor expresses optimism, tempered with moderately valenced realism that acknowledges challenges: "*I quit smoking on Sunday evening. Day 3 today. I feel exhausted, annoyed, bored. But the fight must go on. Keep fighting :)*"

	S-R All_ Before	**S-R All_ After**	**S-R Smoke_ Before**	**S-R Smoke_ After**
Pos. Sent. Intensity	0.0850**	0.0569*	0.0009	-0.0029**
Neg. Sent. Intensity	-0.0010	-0.0031*	0.0002	0.0003

Table 5. Differences in the median values of sentiment variables Before and After cessation for (S)urvivors and (R)elapsers in all tweets (All) and in only smoking-relevant tweets (Smoke). (* p < 0.05, ** p < 0.001, *** p < 0.0001)

Additionally, higher relapse rates are known to correspond to the strength of addiction and severity of withdrawal symptoms [37, 41]. We thus hypothesize that Relapsers' negative affect after cessation is also partly due to their more intense struggles with such negative physical and psychological feelings, which include anger, depression, and anxiety [21], as exemplified in the following tweets by three Relapsers: "*I think my body is busy detoxing, coz I have this constant headache. Must be withdrawel symptoms of the cigarettes.*"; "*Haven't smoked all day my head is killing me* :("; "*pissed off, my back is killing me, my head is swimming with stress and I quit smoking #BadDay*".

Cessation Motives and Setting the Goal to Quit

A long history of research has established that the motivations and orientations with which an individual forms a cessation goal highly influence subsequent commitment, performance, and outcome. Drawing on [27, 35], we perform thematic analysis of our dataset in order to identify the key motivations that trigger an attempt to quit smoking. As explained in our discussion of data preparation, Mechanical Turkers annotated whether or not a smoking-relevant tweet mentions such a motivation. For the tweets that do, we categorize them into 8 classes of "cessation motives". Table 6 presents descriptions and examples of each motive, and Figure 4 shows the proportions of users driven to quit by each. We were able to determine a motive for 224 people.

Motive	**Example Tweet**
Self-Reevaluation: Desire to stop smoking is part of bettering oneself and improving overall life quality	*When did cigarettes do this to my life? Ending the smoking and starting a new me.*
Relationships: Realizes habit's effect on others (e.g., unborn child, family, friends) or is pressured by them to quit	*Today I quit smoking. My son came home with an ashtray he made in art class. FML*
Personal-Threat: Moderate to severe medical symptoms experienced or diagnosed, (e.g. asthma, mouth ulcers, chest pains, heart disease)	*Woke up to serious breathing problems-Mom & friend took me to the hospital for artificial breathing session. Im quitting smoking for good*
Detached-Threat: Becomes fearful for health after exposure to materials about potential dangers or after witnessing other smokers' health complications	*seeing that smoking commercial I legit plan to never have a cigarette again, scared the fuck outta me. Hope I stick with it*
Financial: Cost of cigarettes becomes burdensome or prohibitively expensive	*Just done the budget and smoking is officially unaffordable -- cold turkey here I go*
Cosmetic: Side-effects (e.g., yellow teeth & nails, feeling out of shape) that are less serious than Personal-Threat make smoking unappealing	*never smoking a ciggy again, smell like shit #notworthit*
Holiday: Quitting is a New Year's Resolution or done for a religious holiday	*I'm giving up cigarettes for Lent!!! And forever thereafter!*
Whim: Decision to quit is made casually or on a whim	*Ran out of cigarettes - going to try quitting.*

Table 6. Users' motives to quit smoking and example tweets

We recognize that some motives grow from sincere desires and more permanent circumstances. These motives include Self-Reevaluation, Relationships, and Personal-Threat. Other motives are more casual and fleeting in nature, such as Holiday and Whim. We observe that the former tend to drive Survivors to abstinence (49 Survivors showed sincere motivations, 9 showed casual motivations) while the latter motivate Relapsers (47 sincere, 59 casual), yielding a significant difference between the two groups (χ^2 = 21.67, df=1, p<.001). Figure 4 illustrates these differences, and the following paragraphs explore them more deeply.

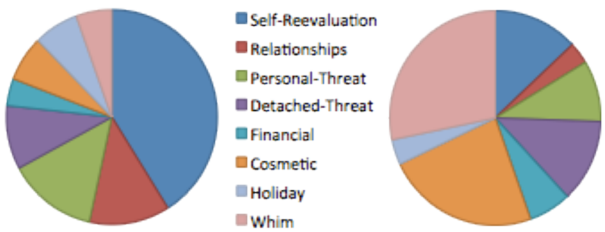

Figure 4. Motives of users who abstain (left) and relapse (right)

First, Self-Reevaluation is a key cognitive step in changing smoking behavior [8], and this internal contemplation produces focus and better chances to meet goals [30]. Motives in our data do show Survivors envisioning their desired state of self as part of such cognitive reconditioning.

Prior work shows that people who recognize the effect of their smoking on others also have better cessation outcomes [34], and we do see Relationships motivating more Survivors than Relapsers. In addition, Personal-Threat motives, such as serious medical concerns, are also more persuasive and see better abstinence rates.

In contrast, we find Relapsers' concerns are more superficial and extrinsically oriented (Cosmetic motives, Detached-Threat), which can induce anxiety and reduce commitment [27]. Whim motives, which overwhelmingly motivate Relapsers, lack any evaluative component and lack accountability. We also see Relapsers combining quitting with more over-the-top and unrealistic fantasies often as part of Holiday resolutions, undermining willpower and causing lack of follow-through [30] (e.g., "*Quit smoking. Exercise. Eat healthier. Start a savings account. Spend less money on pointless shit. Vaycay to Vegas*").

The small numbers available for analysis are due to many users not explicitly mentioning a motive in their tweets. We believe that Holiday and Cosmetic motives are in particular under-reported since investigating when a goal to quit is undertaken (Cessation Date) reveals the pattern in Figure 5. First, we see a spike near the time of New Year's Resolutions. Also noticeable is a peak in late-spring/early summer, which inspection finds is related to users forming goals to get fitter for the summer season via dieting, exercising, and/or quitting smoking (e.g., "*So today is my last day drinking and smoking. Gotta get ma body right fa the summer time and college!*")

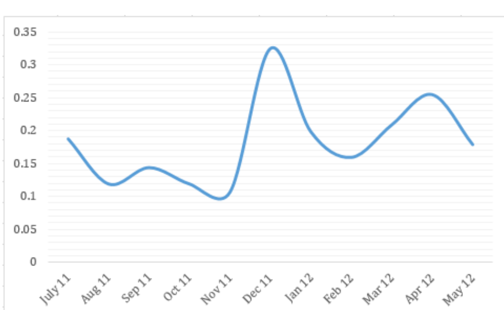

Figure 5. Cessation-event dates throughout the year

The Cessation Process

Stages of Change

Next we analyze whether Survivors and Relapsers demonstrate progression (or lack thereof) through the Transtheoretical Model's stages of change, known as Pre-Contemplation, Contemplation, Preparation, Action, and Maintenance [35]. Essentially, we treat each smoking tweet as an assessment of that user's stage of behavior change. We use the proportion of tweets labeled with a stage as a proxy for the amount of time spent in various stages, which research shows correlates with ultimate outcome [4, 20].

As Figure 6 shows, Relapsers linger more in a preparation phase (Pre-Contemplation, Contemplation, and Preparation combined) than Survivors ($t = 9.63$, $df = 383.30$, $p < .001$).

Figure 6. Proportion of the cessation process spent in each stage of change for Survivors (top) and Relapsers (bottom). Pre-Contemplation, Contemplation, Preparation, Abstinence (Action & Maintenance)

This may be due to Relapsers failing to apply the cognitive, emotional, and evaluative processes necessary to make progress; suffering from chronic procrastination and lack of motivation; and/or having little confidence in their ability to quit -- all of which have been shown to result in longer time spent in pre-abstinence stages and subsequent lower quit rates [15, 34].

We do see evidence of such flawed behaviors in our Relapsers, including procrastination (e.g., "*Quitting red bull today and smoking next week. I might just die of healthy*" and "*Quitting cigarettes cold turkey soon. Going to smoke these 5 more packs of my favorite brands and then be done*"). Relapsers also show low confidence and express pessimism from day one (e.g., "*Trying to quit cigarettes today but dunno if I can do this :-*" and "*Today is my last day smoking…This is going to be hard as hell*"), which contrasts the higher self-efficacy we see from Survivors (e.g., "*Just smoked my last ever cigarette! I know I can quit, yeah baby!*").

Overall, the consistency of our findings with clinical evidence shows that tweeted self-reports about smoking can help to diagnose stage and monitor status on a regular basis.

Cessation Strategies and Treatment Methods

Continuing our analysis of the cessation process, we look more closely at individuals' use of cessation treatments, a variety of which exist and have varying levels of efficacy [16, 26]. Results of thematic analysis to evaluate the cessation strategies employed by users in our dataset are summarized in Table 7 and illustrated in Figure 7.

Strategy	Example Tweet
Cold Turkey; Explicitly mention using no aid	*Hopefully yesterday would be the last time I ever touch a cigarette. Gonna go cold turkey!*
Medication & E-Cigarettes: Nicotine replacement (patch, gum, lozenge) or electronic cig, which gov. considers medicinal	*Ok tweeps, I filled my Chantix prescription today*
Lifestyle Overhaul: Quitting is combined with other positive behavior changes (e.g., with respect to diet, exercise, job)	*got myself a new, well-paid, full time job, quit smoking AND rebuilt my studio setup*
Avoidance & Substitution: Changing routine to avoid tempting situations or replacing cigarettes (e.g., with lollipops)	*Yeah. I've been avoiding smokers to avoid temptation. I'm still afraid I'll cave.*
Group: Quitting with a friend or cessation group	*Today @username and me are quitting smoking*
Self-Help & Alternative Therapies: Hypnosis, books, digital tools	*read "the easy way to stop smoking" by Allen Carr. I'm 6 months cig free with that book*

Table 7. Users' cessation strategies and example tweets

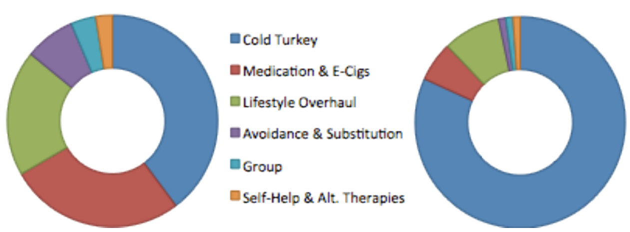

Figure 7. Cessation strategies for smokers who successfully abstain (left) and those who relapse (right)

Clinical research finds that the majority of smokers trying to quit do so without the help of evidence-based cessation treatments, which include over-the-counter and prescription nicotine replacement products, group or individual counseling and therapy, and online programs and self-help plans [11]. Research also finds that smokers trying to quit who use medication or counseling are more successful than those who go unassisted [16]. We see both of these trends holding true in our own dataset. We are able to identify the cessation strategy of 226 people, 156 Relapsers and 70 Survivors. Both Survivors and Relapsers go unassisted more than using any other single method (χ^2 = 15.08, 1 df=1, p<.001). While we see nearly 82% of Relapsers going cold turkey, less than 40% of Survivors do so and instead utilize strategies proven effective more so than Relapsers (χ^2 = 6.90, 1 df=1, p<.01). We also see more Survivors using Avoidance & Substitution, a form of stimulus control important in managing abstinence [8, 34].

DISCUSSION

Theoretical Implications

Our primary goal for this work was to demonstrate that data from social media could bear on the nuanced process of smoking cessation. As mentioned, successful abstinence depends on a stunning range of factors. We separated our Twitter-based measures into four categories: activity, sociality, emotion, and cessation-process.

Our activity measures reflect a strong finding that Relapsers tweet far more than Survivors and are both burstier and more frequent in that tweeting. Combined with the increased likelihood of tweeting at night, these activity patterns suggest Relapsers may be more impulsive, sensation seeking, and challenged by temptation.

Despite this increased tweeting, Relapsers are not more social on a per-tweet basis in terms of friends, followers, or directed communication. This finding fits well with the known role of social support in attaining goals including abstinence from smoking. Furthermore, Relapsers are relatively more negative, aversive, and pessimistic than Survivors in terms of sentiment. As dataset examples show, this sourness may interact with social support (or lack thereof) in that Relapsers tend to appear hostile while Survivors invite and seek connectedness.

Finally, we unravel the cessation process to identify the cessation motives, preparedness, and strategy for about one third of the population studied. While a relatively small number, the results are striking. Relapsers are far more likely to procrastinate before cessation; quit for more casual, shallow, and unrealistic reasons; and chose a cold turkey strategy rather than use effective treatment methods.

Design Implications and Technological Interventions

As mentioned, technological interventions to aid cessation efforts have not been overwhelmingly successful. That our measures differentiate Survivors and Relapsers in ways aligned with theory and experimentation suggest that incorporating social media data may legitimately enhance these interventions. We see two areas for improvement.

First, we can leverage social media to determine if a user is more or less likely to maintain abstinence from day one. For instance, even the single measure of tweeting activity prior to cessation was a strong predictor of relapse as was the amount of social connectivity and the sentiment of posted content, all of which can provide interventions such as mobile phone applications with high level direction as to the level of assistance a user needs.

Second, interventions can be tailored using the different motives, attitudes, and behaviors captured about its users. An application that provides information and reminders on the smoking cessation goal might base its content, timing, and audience on these social media based measures. As examples, users going cold turkey and tweeting frequently at night can receive an intervention designed for susceptibility to temptation, strong addiction levels, and end-of-day support. Users relatively lower in social interaction can be encouraged to reach out or can have support-seeking messages automatically posted to their feed. Those whose content signals poor mood can receive advice to reflect on positive aspects of abstinence or to broadcast more optimistic updates about their cessation attempt. Users whose cessation begins on a holiday can be guided through self re-evaluation and encouraged to consider substantive consequences of their smoking habit.

Lastly, certain activities and changes in behavior can serve as relapse warning flags. For example, context-aware applications could detect a visit to a bar and provide temptation support. A shift to noticeably higher posting activity, increasingly negative affect, a reduction in active social ties, or a lingering in certain behavior change stages can trigger a tool to intervene with therapeutic recommendations or with alerts to trusted connections.

We envision a largely, but not fully, automated intervention system that combines social media with user-provided data. User input can help fill gaps, provide labels on the social media data, and set cessation and intervention preferences and boundaries. A user's explicit contribution of detail about smoking-related behaviors and mindsets can also address activities and thoughts that may occur without a corresponding post on Twitter, which would help mitigate issues of invisible or ambiguous data and could play a role in a more holistic cessation treatment overseen by a clinician, who traditionally struggles with ensuring a patient fully reports behavior.

Finally, we recognize that for smoking cessation, individual definitions of what constitutes success can vary. One person may strive for total, permanent abstinence, whereas another may just want to cut back. Deducing and accommodating such differences in expectations will be key when personalizing intervention tools.

Future Work

Despite its array of benefits, there are limitations to relying on Twitter data that we acknowledge to avoid making unfounded claims and to suggest compensating future plans. By focusing on Twitter users, we are drawing conclusions on people who have chosen to broadcast their smoking habits, desire to quit, and positive and negative progress towards that goal. There could be fundamental differences between these people and those who undertake goals more privately, in different online settings, or in some other manner to which we do not have access. Of course, this challenge is seen in clinical interventions as well, where participants often represent a minority of smokers [16]. Thus though we see agreement with CDC statistics that suggest our sample is aligned with the full population of smokers, an essential next step is to extend our study with data from additional online and offline sources to see how much our findings translate to broader scenarios.

Within the Twitter sample, there are potential confounds that are difficult to eliminate. For instance, age is not available in the social media data and may explain some differences observed between Survivors and Relapsers. In practice, age and other demographic details, including smoking specific information like years of smoking, may need to be entered by the user in a technological intervention. Also related to sampling, we are limited by our 2-year time window; and although this is similar to traditional clinical settings in which patients typically stop assessments after 24 months, expanding the scope of data to longer time periods is an opportunity for additional insight into complications like long-term relapse.

CONCLUSION

In this paper, we leveraged socio-technical systems to study human behavior and inform next-generation technology designs, demonstrating the immense opportunity offered by social media to better understand, model, and predict why and how the smoking cessation behavior change process happens. We combined automatic and crowdsourced techniques to detect Twitter users trying to quit smoking; analyzed their network data, tweets, and smoking-relevant content; and explored the cessation process in the year leading up to the cessation-event as well as the year after. We illustrated that this type of data closely approximates real world data yet is much less cumbersome to obtain.

We also exhibited how the Twitter medium captures emotional expression, mediates social interaction, documents behavior, and reflects behavior change, finding clear distinctions between individuals who do and do not remain abstinent over the observation period according to such features as posting volume and frequency, social subscription and outreach, and affective intensity. Our contributions validate and advance prior work on what fine-grained aspects of these features impact and predict success and failure during smoking-related behavior change.

Finally, we provided design implications to drive forward and steps for developing systems that interpret these signals as part of tailored cessation support and intervention.

ACKNOWLEDGEMENTS

We gratefully acknowledge guidance from Meredith Ringel Morris, Munmun De Choudhury, and other members of the Nexus research group and Microsoft Research Redmond.

REFERENCES

1. Abroms, L., Padmanabhan, N., Thaweethai, L., Phillips, T. iPhone Apps for Smoking Cessation: A Content Analysis. *American Journal of Preventive Medicine* (2011), 279-285.
2. Alderman, M. Goals and goal setting. *Motivation for achievement* (1999), 88-111.
3. American Cancer Society. Cancer Facts and Figures. (2012).
4. Aveyard, P., Massey, L., Parsons, A., Manaseki, S., Griffin, C. The effect of Transtheoretical Model based

interventions on smoking cessation. *Social Science & Medicine 68*,3 (2009), 397-403.
5. Bahr, D. B., Browning, R. C., Wyatt, H. R., & Hill, J. O. Exploiting social networks to mitigate the obesity epidemic. *Obesity 17*,4 (2009), 723-728.
6. Bandura, A. Self-efficacy: The exercise of control. (1997).
7. Bandura, A. & Locke, E. A. Negative self-efficacy and goal effects revisited. *Applied psychology 88*,1 (2003), 87.
8. Borland, R., Partos, T., Yong, H., Cummings, K., Hyland, A. How much unsuccessful quitting activity is going on among adult smokers? *Addiction 107* (2012), 673-682.
9. Bradley, M. M. & Lang, P. J. Affective norms for English words (ANEW). *Technical Report C-1* (1999).
10. Campbell, R. S. & Pennebaker, J. W. The secret life of pronouns: flexibility in writing style and physical health. *Psychological Science 14*,1 (2003), 60-65.
11. Centers for Disease Control. Current Cigarette Smoking Among Adults - United States. (2011).
12. Chase, C. Public Policy Statement on Nicotine Dependence and Tobacco. *American Society of Addiction Medicine* (2010).
13. De Choudhury, M., Counts, S., & Horvitz, E. Major life changes and behavioral markers in social media: case of childbirth. In *Proc. CSCW* (2013), 1431-1442.
14. DiClemente, C. C., Fairhurst, S. K., & Piotrowski, N. A. Self-efficacy and addictive behaviors. (1995), 109-141.
15. DiClemente, C. C., Prochaska, J. O., Fairhurst, S. K., Velicer, W. F., Velasquez, M., & Rossi, J. The process of smoking cessation: an analysis of precontemplation, contemplation, and preparation stages of change. *Journal of consulting and clinical psychology 59*,2 (1991), 295.
16. Fiore, M. Treating tobacco use and dependence: Clinical practice guideline. (2008).
17. Fowler, J. H. & Christakis, N. A. Dynamic spread of happiness in a large social network. *BMJ: British medical journal 337* (2008).
18. Gulliver, S. B., Hughes, J. R., Solomon, L. J., & Dey, A. N. An investigation of self-efficacy, partner support and daily stresses as predictors of relapse to smoking in self-quitters. *Addiction 90*,6 (1995), 767-772.
19. Heatherton, T., Kozlowski, L., Frecker, R., Fagerstrom, K. The Fagerstrom test for nicotine dependence. *British journal of addiction 86*,9 (1991), 1119-1127.
20. Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. The dose-effect relationship in psychotherapy. *American Psychologist 41*,2 (1986), 159.
21. Hughes, J., Stead, L., Lancaster, T. Antidepressants for smoking cessation. *Cochrane Database Syst Rev* 1.1 (2007).
22. Hunt, W. A. & Matarazzo, J. D. Habit mechanisms in smoking. *Learning mechanisms in smoking* (1970), 65-90.
23. Hutto, C., Yardi, S., Gilbert, E. A longitudinal study of follow predictors on twitter. In *Proc. CHI* (2013), 821-830.
24. Jamison-Powell, S., Linehan, C., Daley, L., Garbett, A., & Lawson, S. I can't get no sleep: discussing #insomnia on twitter. In *Proc. CHI* (2012), 1501-1510.
25. Kouloumpis, E., Wilson, T., Moore, J. Twitter sentiment analysis: The Good the Bad and the OMG! In *ICWSM* (2011).
26. Lemmens, V., Oenema, A., Knut, I. K., & Brug, J. Effectiveness of smoking cessation interventions among adults: a systematic review of reviews. *EJCP* (2008).
27. Locke, E. A. & Latham, G. P. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist 57*,9 (2002), 705.
28. Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. Goal setting and task performance: 1969-1980. *Psychological bulletin 90*,1 (1981), 125.
29. Manchón Walsh, P., Carrillo, P., Flores, G., Masuet, C., Morchon, S., Ramon, J. Effects of partner smoking status & gender on long term abstinence rates of patients receiving smoking cessation treatment. *Addictive behaviors* (2007).
30. Oettingen, G., Pak, H., Schnetter, K. Self-regulation of goal-setting. *Personality & social psych. 80*,5 (2001), 736.
31. Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. The development and psychometric properties of LIWC2007. (2007).
32. Pennebaker, J. W., Francis, M. E., & Booth, R. J. Linguistic inquiry and word count: LIWC 2001. (2001), 71.
33. Prochaska, J. J., Pechmann, C., Kim, R., & Leonhardt, J. M. Twitter=quitter? An analysis of Twitter quit smoking social networks. *Tobacco control 21*,4 (2012), 447-449.
34. Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. In search of how people change: Applications to addictive behaviors. *Journal of Addictions Nursing 5*,1 (1993), 2-16.
35. Prochaska, J. O. & Velicer, W. F. The transtheoretical model of health behavior change. *American journal of health promotion 12*,1 (1997), 38-48.
36. Prochaska, J. O., Velicer, W. F., DiClemente, C. C., & Fava, J. Measuring processes of change: applications to the cessation of smoking. *Journal of consulting and clinical psychology 56*,4 (1988), 520.
37. Redding, C. A., Prochaska, J. O., Paiva, A., Rossi, J. S., Velicer, W., Blissmer, B. J., Greene, G. W., Robbins, M. L., & Sun, X. Baseline stage, severity, and effort effects differentiate stable smokers from maintainers and relapsers. *Substance use & misuse 46*,13 (2011), 1664-1674.
38. Robinson, L. M. & Vail, S. R. An Integrative Review of Adolescent Smoking Cessation Using the Transtheoretical Model of Change. *Journal of Pediatric Health Care 26*,5 (2012), 336-345.
39. Rude, S., Gortner, E.-M., & Pennebaker, J. Language use of depressed and depression-vulnerable college students. *Cognition & Emotion 18*,8 (2004), 1121-1133.
40. Teodoro, R. & Naaman, M. Fitter with Twitter: Understanding Personal Health and Fitness Activity in Social Media. In *ICWSM* (2013).
41. Velicer, W. F., Redding, C. A., Sun, X., & Prochaska, J. O. Demographic variables, smoking variables, and outcome across five studies. *Health Psychology 26*,3 (2007), 278.