Do Group Memberships Online Protect Addicts in Recovery Against Relapse? Testing the Social Identity Model of Recovery in the Online World

ELAHE NASERIAN and MIRIAM KOSCHATE, Department of Psychology, University of Exeter, United Kingdom

The Social Identity Model of Recovery (SIMOR) suggests that addiction recovery is a journey through time where membership in various groups facilitates success. With the help of computational approaches, we now have access to new resources to study whether a wide variety of different online communities can be part of the addiction recovery journey. In this work, we study the effects of two main social factors on recovery success: first, multiple group membership defined in terms of richness of online community engagement; second, active participation operationalized as the evenness in engagement with these groups. We then model recovery from addiction by applying the extended Cox regression model which accounts for the effect of these two factors on time to relapse. We applied our analysis to a dataset of 457 recovering opioid addicts that self-announced the date of their recovery, indicating that at least 219 (48%) addicts relapsed during the recovery period. We find that multiple group membership – in terms of the number of other forums that a subject had posted in - as well as active participation - in terms of how evenly their posts were spread amongst the different forums - reduced the risk of relapse. We discuss our findings with regards to the opportunity, but also risk, that online group membership poses for recovering opioid addicts, as well as the possible contribution that computational social science methods can make to the study of addiction and recovery.

68

CCS Concepts: • **Applied computing → Law, social and behavioral sciences**; **Psychology**;

Additional Key Words and Phrases: addiction recovery, social cure, online communities, survival analysis

ACM Reference Format:
Elahe Naserian and Miriam Koschate. 2021. Do Group Memberships Online Protect Addicts in Recovery Against Relapse? Testing the Social Identity Model of Recovery in the Online World. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 68 (April 2021), 18 pages. https://doi.org/10.1145/3449142

1 INTRODUCTION

Opioid addiction presents one of the most pressing public health issues of the day and has been declared a national health emergency by the U.S. government in 2017 [47], with an average of 128 overdose deaths from opioids every single day in the U.S. alone [9]. Worldwide, 118,000 deaths in 2015 were directly associated with an opioid use disorder [52].

Treatment options for opioid addiction vary widely, from community-based treatments as outpatients without medication to medication-assisted inpatient treatment, and several other options in-between [21]. However, relapse rates within 6 months of treatment remain high (38 − 77%; [35]). Although several relapse episodes are likely before abstinence can be maintained, those who

Authors' address: Elahe Naserian, e.naserianhanzaei@exeter.ac.uk; Miriam Koschate, m.koschate-reis@exeter.ac.uk, Department of Psychology, University of Exeter, Exeter, United Kingdom.

2573-0142/2021/April-ART68 $15.00
https://doi.org/10.1145/3449142

can maintain abstinence for longer periods significantly lower their risk of a subsequent relapse [11].

A prominent social psychological model in the area of addiction research - the Social Identity Model of Recovery (SIMOR; [4]) - suggests that meaningful membership in various social groups underpins the successful transition from addiction to long-term recovery. SIMOR builds on a social identity approach to health and well-being, known as the 'social cure' [19]. Within this research tradition, group membership is seen as part of an individual's self-concept - their social identity - once the individual sees themselves, and is seen by others, to belong to a particular social group or category [26].

Much of the current literature on the role of social groups in addiction recovery focuses on an individual's offline support network, including family, friendship groups, peer support groups such as Narcotics Anonymous (NA), and groups related to a person's interests and hobbies [19]. More recently, research has found that, in addition to the positive effects of offline networks, online recovery support groups can promote addiction recovery [6]. However, no study to date has investigated the wider effect of online group memberships beyond online recovery groups, that is, the 'social cure' effect of having a variety of online social identities.

With the prevalence of online social networks, a large and increasing number of people are going online for information, advice, debate, and support, and those with substance-use problems are no exception. Several of the largest online community websites include forums dedicated to the topic of opioid addiction and recovery from opioids. These online communities advertise themselves to substance users who seek advice, support from a community, or simply wish to share their addiction stories, while allowing them to stay anonymous.

The public and anonymous nature of these online forums allows us to study addiction and recovery in new ways by providing longitudinal data of subjects that may be hard to recruit for traditional surveys or interview studies. The forum platform, Reddit, additionally allows us to examine how membership in a large variety of different online communities affects recovery from opioid addiction. Reddit hosts well over 100,000 different sub-forums ("sub-reddits") where content is shared within a community of interest, for instance about politics, business, parenting, medical conditions, sports, literature, music, video games, and life choices. Some of these communities are close-knit, with restrictions on who is allowed to post, whereas others are open to contributions from anyone registered on Reddit.

By examining activity in "sub-reddit" communities, we study the effects of online social identities on opioid recovery by going beyond online recovery support groups. According to SIMOR [4], addiction recovery is a journey through time where membership in *various* groups facilitates success. Specifically, we study the effect of two main social factors on success of recovery: first, multiple group membership which is operationalized in terms of richness of group membership. Second, active participation which is measured by considering the evenness of engagement with these groups. We then model recovery from addiction by applying survival analysis which accounts for the effect of these two factors on time to first relapse. We make the following contributions:

(1) We first identify drug addicts on the online platfrom Reddit whose recovery progress can be traced. The subjects are then categorized into relapse and non-relapse cohorts based on whether relapse has been self-reported or inconsistent recovery progress can be identified from posts. The final dataset consists of 457 subjects who have self-announced their recovery, including 219 (48%) subjects who have relapsed.

(2) We estimate relapse over time in our dataset of Reddit users [1], applying a model based on survival analysis. We observe that the sharpest decrease in the survival probability (probability of

[1]by 'users', we mean 'the social network users', unless we specifically mention the 'substance users'

not experiencing the first reported relapse) happens during the first 6 months of recovery, and the probability of relapse decreases with increasing time. The median time to relapse is 12-13 months, in other words, during the first year, almost 50% of the population have relapsed.

(3) A social identity approach is used to better understand the impact of online communities. In this regard, we study two measures based on SIMOR that are likely to affect the success of recovery: multiple group membership and active participation. The strongest predictor for relapse in our cohort is multiple group membership, that is, the number of Reddit communities a recovering addict is engaging with. Recovering addicts who increase their group membership by one group, decrease their chance of relapse by 4%. In addition to the number of groups, it also appears to be important to actively participate in them equally. More evenly engaging in different groups significantly reduces the risk of relapse.

(4) We discuss our findings with regards to the opportunity, but also potential risk, that online group memberships pose for recovering addicts, as well as the possible contribution that computational social science methods can make to the study of addiction and recovery. We discuss the benefits that online group memberships can provide as a potential alternative to face-to-face groups in making addiction recovery sustainable as well as the need for future research to better understand the factors that allow addicts in recovery to build (online) social capital.

2 BACKGROUND AND MOTIVATION

Online peer support communities have been shown to have a positive impact on recovery from different illnesses, by increasing self-empowerment and by helping with stress, depression, and encouraging positive coping behavior [20], [43], [16]. Similarly, online peer support groups for addiction recovery have been shown to have a positive impact[6]. However, it is currently unclear whether offline 'social cure' effects - that is, the impact of being a member of a variety of (unrelated) groups - also holds for online community memberships. To the best of our knowledge, our work is the first to investigate the influence of multiple online group memberships on addiction recovery.

Understanding the impact of online community memberships, beyond the direct effect of online recovery groups, may help us tailor support for those with limited access to offline social networks. Access to a supportive face-to-face network of peers can be limited, for instance due to geographical isolation, mobility issues or significant caring responsibilities. Furthermore, according to the World Health Organisation ([53]), drug use is the most stigmatised health condition in the world which may make it harder for individuals to attend face-to-face support group meetings [7] and join groups of people with similar interests. The worldwide lockdown during the Covid-19 pandemic has additionally raised concerns for those recovering from addiction by severely limiting face-to-face interactions, not only with support groups but also the wider support network of family and friends, among other factors [50]. Online communities may therefore play an increasingly important part in providing social support and information [38], [40].

The online platform Reddit provides a unique way to analyse the various aspects of group participation on the sustainability of opioid recovery and chance of relapse over time. It also provides a first indication of the role that anonymous online groups (rather than face-to-face social groups and online recovery groups) can play in supporting recovery. In this section, we briefly present studies that provide support for SIMOR in offline contexts. We will also briefly review the use of social media to understand addiction more widely, and the impact of online recovery groups specifically.

2.1 Social Groups and Recovery from Addiction

SIMOR suggests that social group memberships are a key part of making a recovery from addiction sustainable [4]. Being part of a social group may provide (1) a sense of connectedness with relevant others, (2) meaning, purpose, and worth as a member of a positively valued group, (3) social support in the form of psychological and physical resources, and (3) control, efficacy, and power [19]. Studies on the 'social cure' usually ask survey respondents to report the number of meaningful groups that they are part of [2]. Findings show that a higher number of self-reported group memberships are associated with generally better health and well-being, and lower levels of smoking and drinking [39].

Further evidence from the field of addiction highlights the important role of social groups in recovery [5]. For example, in [2], the authors found over a 6-month follow-up self-report study that residents of a therapeutic community showed marked shifts in their social identity. Those who increased the proportion of group memberships with non-using groups reported a decreased use of substances, and an increased sense of a recovery identity rather than a user identity. The overall reported number of meaningful group memberships was low, with a median of 4 group memberships.

Joining and participating in traditional support groups, such as Alcoholics Anonymous (AA) or Narcotics Anonymous (NA), has also shown significant promise in assisting substance users in maintaining their abstinence [15], [32], by providing an encouraging and supportive community and by facilitating programs for addiction management and recovery. The effectiveness of such groups is highly dependent on regular attendance [25], suggesting that active participation is an important aspect of group membership.

2.2 Social Media and Addiction

Increasingly, researchers use computational social science techniques to gain a better understanding of addiction and recovery. To do so, researchers make use of naturally occurring data from various platforms including Twitter, Facebook and relevant online forums. Such data can provide insights into addiction and recovery that is unencumbered by researcher expectations on subjects (e.g., social desirability, demand characteristics). Although online data often lack demographic and context information on subjects, a range of activity variables (e.g., likes, number of posts/comments) and linguistic variables (e.g., emotion words, tags) are used to predict the phase of recovery that an individual is likely to reach.

Several studies exist that have examined addiction and recovery through the analysis of online data. However, these studies often focus on recovery as a process of personal transformation. For example, [30], [29], [17] model the recovery and relapse by defining linguistic and content features of posts to understand in which recovery phase an individual is, and how amenable they might be to recovery. Although studying the characteristics of those "in recovery", or who regard themselves as "recovered", helps us to understand the personal aspects of the problem, it fails to identify the mechanism of change, or the social context in which change occurs [4].

Focusing on the beginning of the recovery journey, the authors in [29] train a binary classifier using linguistic features of posts to predict an addict's transition from casual drug discussion forums to drug recovery forums. Similarly, in [17], authors built a computational model to predict an addict's propensity for seeking drug recovery interventions. They developed a model to predict whether an addict is going to contribute to a recovery sub-forum on Reddit, based on their activity in other sub-forums.

To identify several addiction phases from (recovering) addict's posts, the authors of [30] analyzed the process of drug withdrawal, recovery and relapse on Forum77, MedHelp.org[2]. They trained a sequence classifier taking several activity and linguistic features into account. Similarly, for smoking cessation, the authors of [34] show with activity variables, social variables and emotional

[2]an online health forum for substance abuse recovery

variables extracted from Twitter that those who succeed with recovery (quit smoking) behave differently than those who relapse.

There are a few studies which have investigated the impact of recovery specific online communities on recovery [42], [18]. For instance, authors in [1] analysed the data of subjects from Sober Grid, a recovery social network site - smart phone application, to identify the demographics, engagement patterns, and recovery outcomes of active subjects and to examine between generational group differences on activity variables and recovery outcomes. In [3], authors carried out qualitative analysis on data from IntheRooms, an online addiction recovery tool, to study the effectiveness of this online recovery community. In [6], the authors investigated whether increased levels of social interactions and developing a positive identity can predict the retention in an online recovery program. They conducted their analysis on a Facebook page dedicated to support addicts who are in their early recovery and measured the level of participation by number of received 'likes' and increased use of the word 'we'. In contrast to these studies, we do not look at the participation in recovery-specific social network sites but instead we examine the effect of wider online community membership, largely consisting of online communities unrelated to addiction/recovery.

Our study builds on the insights of these studies by using social variables to understand protective factors against relapse from addiction. We will look at activity in several sub-forums on Reddit similar to [17] but aim to predict the likelihood of relapse rather than the propensity for moving to a recovery forum. Similar to [6], we will use SIMOR to understand how social group membership shapes the recovery journey. But rather than focusing on a single recovery forum and retention in this forum, we will use activity in a large number of sub-forums, many of which are unrelated to substance-use or recovery, on the same platform to look at the likelihood of relapse. Our research also differs from [30] and [34], who also look at recovery versus relapse, by looking at several group memberships rather than interpersonal social relations as in [34] or a particular recovery forum as in [30].

3 DATA

The most important challenge in using online data to understand clinical or social phenomena is to establish validity across all the components of study, from collecting the dataset to choosing the method, and selecting good quality measures [10]. Where validity is not established, no conclusions for theory or practice can be drawn. In the following, we provide a detailed description of the ways in which we have sought to establish validity through careful corpus building (identifying positive and negative samples), algorithm selection, and feature selection.

Choice of online platform: Our study uses data from Reddit, a public social news and discussion website where user-created content is organized into topic-based boards called "subreddits". Reddit accommodates thousands of different subreddits which cover a variety of subjects such as news, politics, science, movies, video games, music, etc. Reddit provides a platform for individuals to be part of several communities based on their interests or hobbies. As this study aims to study the effect of multiple online group memberships on the likelihood of a relapse, Reddit is a platform that affords this opportunity by providing access to individuals' data outside addiction/recovery-specific forums.

Acquiring the ground truth: In the absence of clinical assessments, researchers look for alternative signals to identify positive and negative instances of health behavior. Self-report or self-disclosure is introduced as one of the robust strategies for acquiring the ground truth about different health outcomes [10]. It is especially attractive in the case of problems where social stigma might discourage face-to-face reports, such as mental health problems and addiction [36]. For example, in [46], [45] authors used self-reported information in two online communities in Reddit (r/stopsmoking and r/stopdrinking) as a reliable, high-quality signal to identify and trace abstinence from smoking

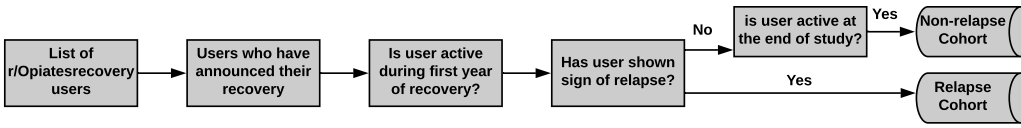

Fig. 1. Diagram of data collection and preparation for identifying "Relapse" and "Non-relapse" cohorts.

and alcohol drinking. In [54], authors identified subjects attending AA meetings through the study of self-report on Twitter. In addition to addiction and recovery, self-report has been widely used for identifying individuals who suffer from mental health disorders, such as post-traumatic stress disorder, schizophrenia, or depression [12], [13], [49]. We use 'self-disclosure' in the form of users posting about their recovery/relapse status as signals to identify the subjects who have started their recovery and those who have, unfortunately, relapsed. We combine self-disclosure with 'human assessment' to ensure that correct information is extracted from self-reports. Specifically, a researcher with psychology training manually went through all inconsistent recovery announcements to assess whether the user has relapsed (see Phase II in Preparing and Building the Cohorts for detailed information).

Ethics, Privacy, and Disclosure: Before work commenced, the study received ethical approval from the ethics committee of the Department of Psychology at the University of Exeter. This paper uses publicly accessible Reddit data for analysis. Reddit user-names were used to collect quantitative data across different sub-reddits (see Data section for detailed information). Once these numbers were collected, the user ID was replaced with completely anonymous subject numbers in the working dataset. No post content is included in the working dataset which consists of only numerical variables. We only use paraphrased expressions rather than original quotes to illustrate our NLP methods in the Data section.

3.1 Preparing and Building the Cohorts

Figure 1 illustrates the four steps involved in our data collection and preparation process towards identifying two cohorts: a relapse cohort and a non-relapse cohort. In phase I, we collect the data from r/OpiatesRecovery, one of the most popular addiction recovery subreddits. In phase II, we identify subjects who have reported their recovery, and among them, detect those who have shown signs of relapse. This gives us two sets of subjects: non-relapsed subjects and relapsed subjects. In phase III, we collect the reddit content for each subject, and filter out those subjects where only insufficient information can be obtained. In phase IV, we build the non-relapse cohort by including the data from recovery start date until last announcement for each non-relapsed subject, and build the relapse cohort by including the data from recovery start date until relapse date for each relapsed subject.

Phase I: Collecting and Preparing Datasets

For the initial data collection step, we searched for popular subreddits related to recovery from drug addiction. After a few rounds of discussion and randomly reading comments and posts from several subreddits, we chose r/OpiatesRecovery. This subreddit is used by a group of people who are supporting each other through the journey of recovery from opioid addiction. We collected the whole history of this subreddit between February 2012 and June 2019. After cleaning the data from adversarial content, bots, etc., the dataset contains 295,232 posts and comments from 18,125 users.

Phase II: Identifying the Qualified Subjects for Relapse and Non-relapse Cohorts.

Subjects who have started their recovery: To identify qualified subjects for our analysis, we first manually examine posts and comments containing some general recovery tags, for example "clean", "off", "in recovery". Among these posts, we then look for statements used to announce recovery progress. Based on a snowball sampling approach during this inspection phase, we obtain a list of statements which are commonly used for reporting on recovery progress. Below are some (paraphrased) examples:

- "I've been off drugs for over 2 months."
- "I'm 21 days clean."
- "It is almost a year now that I'm in recovery."

Several regular expression patterns then are designed to look for these statements in our dataset. Through this process, we find 2,950 instances of recovery being announced by 1,651 subjects. Some of these subjects report their recovery progress more than once which helps us to examine their consistency in recovery. Notably, not all of these subjects are qualified for our final analysis. There are some further steps for preparing the relapse and non-relapse cohorts which will be explained in the following.

Subjects who have relapsed: Among those who begin their recovery journey, a significant number relapse. To identify them, we follow the same procedure as for identifying recovery: we start with some general tags (such as "relapse", "using again", "in use") to look for statements commonly used to report relapse. We find several statements, a few are listed bellow (paraphrased):

- "... I relapsed yesterday."
- "... a couple of days ago I started using again."
- "... two months ago I used."

By designing several regular expression patterns, we then search for relapse reporting among all the posts and comments from those who have announced their recovery progress. We find 1,069 posts and comments from 370 subjects mentioning relapse.

Some of the subjects do not self-announce their relapse directly but show an inconsistency in their recovery progress. Consider the case of a user who initially declares two months of recovery. However, a few months later, they announce that they have been clean for one month. This inconsistency in recovery progress indicates a relapse between two reports of recovery. By analysing the time announcements of subjects in the recovery forum, and noting a lack of progress in some of them, we identify 93 subjects who likely relapsed between two announcements of recovery. A researcher with psychology background manually went through these inconsistent recovery announcements, marking (yes/no) whether the user indeed relapsed based on information from their posts. Out of 93 subjects with inconsistent progress, 51 are added to the list of relapsed subjects.

It should be noted that if there are multiple recovery journeys for a user (starting, and then relapsing and starting over), we consider the first recovery/relapse journey for the analysis.

Phase III: Obtaining the Historical Data of Qualified Subjects.

We start with our 1,651 subjects and inspect their historical data from all Reddit sub-forums (i.e., all the subreddits they had posted in). Since we wish to study the activity of addicts in recovery from the beginning of their recovery, subjects whose self-announced recovery time exceeds their time on Reddit by more than a year are excluded from analyses. For instance, if a subject self-announces their recovery start for the year 2010, but the first Reddit post dates from 2012, they are not included in our dataset. To allow sufficient information about each individual subject in our dataset, we only retain subjects who posted at least 10 posts and comments for a minimum of three months during

Table 1. Summary Statistics of “Relapse” and “Non-relapse” cohorts.

	All	Non-relapse	Relapse
Total subjects	457	238	219
Total Posts	237,298	137,030	100,268
Total Subreddits	4,582	3,513	2,268
Median Posts / subjects	158	157	160
Median Subreddits / subjects	14	18	11

the first year of recovery across all Reddit sub-forums [3]. We do not, however, impose a restriction on the minimum number of posts for any individual subreddit. This leaves us with 1,081 active subjects, of which 335 are identified as subjects who relapsed and 746 as non-relapsed subjects.

Phase IV: Building the Relapse and Non-relapse Cohorts.

Building the non-relapse cohort : Among the subjects who did not relapse, all those who are still active in the r/OpiatesRecovery subreddit by the end of our data collection in 2019 are retained in the dataset. This leaves us with 238 subjects in the non-relapse cohort. For each of these subjects, we include their Reddit history from recovery start date until the last recovery announcement in our dataset.

The reason for this extra post-processing is to ensure the quality of our data by including only subjects in the non-relapse cohort who have continuous data until the end of our data collection period [4].

Building the relapse cohort: From the subjects who have shown signs of relapse - by mentioning it themselves or based on inconsistent recovery progress - we select those for whom the date of their first relapse can be estimated at the scale of a month. For the subjects who have shown inconsistent recovery progress, we can identify the time interval during which the subject might have relapsed (which is between the last report of the first recovery and start date of the next recovery period). We exclude those subjects from our dataset for whom we cannot estimate the month in which the relapse occurred. We also exclude those who have not posted at all on Reddit between their recovery start date and relapse date and therefore have no valid data for our predictor measures. In total, we include 219 subjects in the relapse cohort. For each subject, we include their Reddit history from recovery start date until their relapse date in our dataset.

We take further steps to manage issues of data bias and ensuring the quality of our dataset by removing adversarial content, bots, and posts from deleted accounts, and by setting minimum engagement thresholds (minimum number of posts and minimum active months). Table 1 gives summary statistics of the data in the relapse and non-relapse cohorts. It should be noted that the total subreddits refer to the total number of *unique* subreddits a subject is a part of.

4 METHODS

4.1 Models for Time-to-Event Analysis

Time-to-event analysis is used when studying the occurrence of an event among a population. The event of interest in our case is the relapse to opioid use after recovery began. The primary goal of

[3] To ensure we have sufficiently reliable data, we have excluded the lower quartile (< 3 month of activity; <10 posts) of subjects. Importantly, the minimum of 10 posts is across the whole of Reddit, not specifically for the forum OpiatesRecovery or any other subreddit individually.

[4] By keeping the subjects who are still active in the subreddit till the end of study, we make sure of "non-informative" censoring. We will further explain this in Section 4.1

time-to-event analysis is to understand factors which affect the probability of experiencing/not-experiencing the event of interest over a limited study period. The time until a subject experiences the event is the 'survival time'. However, not all the subjects experience the event by the time the study ends. Observations of subjects that do not experience the event of interest are called censored data. Traditional regression methods are not equipped to handle censored observations. These data are either ignored or the observed survival time is equated with the unobserved total survival time, with both methods leading to biased results [41].

Survival analysis is one type of time-to-event analysis which has been widely adopted when the research interest is a combination of *whether* the event has occurred (binary outcome) and *when* it has occurred (continuous outcome). It provides unbiased survival estimates by utilizing the information provided by subjects who have experienced the event, as well as those who have not (censored data). Before applying survival analysis to our research problem, we need to define a few terms:

Relapse Event: Our event of interest is "relapse". A user is said to have experienced the relapse event if they have explicitly reported it, or implicitly shown it by having inconsistent recovery progress.

Survival Time: We measure the survival time in months. For subjects in the non-relapse cohort, who never experienced the relapse event, survival time is from start date of recovery until their last recovery progress report. For subjects in our relapse cohort, survival time is measured from recovery start date until the month of relapse.

Survival Function: The survival function gives the probability that a person survives longer than some specified time t.

Hazard function: The hazard function at a time t specifies the instantaneous rate at which subjects experience the relapse, given that they have survived up to time t.

Censoring: Censoring occurs when we have some information about individual survival time, but the exact survival time is unknown. There are different types of censoring, of which right censoring is the most common. Right censoring occurs when a subject leaves the study before an event occurs ("loss-to-follow-up"), or the study ends before the event has occurred ("end-of-study"). Censoring in survival analysis should be "non-informative" - subjects who drop out of the study should do so due to reasons unrelated to the study [37]. Informative censoring occurs when subjects are lost to follow-up due to reasons related to the study, which can increase the risk of bias and result in lower accuracy. In our study, we have both types of right censoring. However, in case of loss-to-follow-up censored data, we do not know the exact details of why we lost the subject and if the reason is related to the outcome (e.g. the subject may no longer contribute to the subreddit because they have started using drugs again). Therefore, we only include subjects who are censored at the end of the study (that is, those who have not relapsed and are still active in the r/OpiateRecovery subreddit at the end of the data collection period).

4.2 Survival Analysis Techniques

We employed three different survival analysis techniques. To determine the rate of relapse, we used Kaplan-Meir estimator [22]. When the underlying data is censored, this model provides an estimation of the survival function. It estimates the probability of not relapsing as a function of time. We also construct the life table applying actuarial estimation. This estimator takes into account that we do not exactly know when an event occurs during each time interval. It calculates the conditional probability of relapsing within the interval, given no relapse at the beginning of it. The main difference with the Kaplan-Meir estimator is that censoring is assumed to occur, on average, halfway through the interval.

Table 2. Cumulative probability of surviving and hazard at different time intervals.

Time-interval (months)	Survival probability	pdf	Hazard
0-6	1.0000	0.0665	0.0797
6-12	0.667	0.0239	0.0401
12-24	0.523	0.0125	0.0279
24-48	0.373	0.0064	0.0216
>48	0.219	NA	NA

In order to explore the effect of various factors on time to relapse, we employed an extended Cox model [27]. This survival analysis regression method explores the relationship between the event of interest and the factors that affect the time at which the event occurs. It allows us to study how survival probabilities (or hazard rates) change with changes in the studied factors. Unlike the basic Cox model, the extended version is designed to accommodate time-dependant variables, that is, those variables whose value for a given subject may differ over time. Cox modeling does not make any assumption about the statistical distribution of the survival times, unlike most other statistical models, which makes it an appropriate choice for our research problem.

5 MEASURES

To choose measures for analysing the effect of online group memberships on recovery success, we follow the social identity model of recovery (SIMOR). According to SIMOR, multiple social group memberships, and actively participating in different groups, helps to sustain recovery from addiction [4]. To consider both aspects of group membership - number of group memberships and participation - we measure the diversity of a subject's group membership. A diversity index is a quantitative measure that reflects how many different types (of groups in our case) there are in a dataset and simultaneously takes the distribution of individuals across those types into account. Diversity consists of two components: richness and evenness.

Richness, or the number of different types in a dataset, is the simplest metric used to indicate diversity [51] and is applied frequently in research [31]. For each subject, we calculate the richness of group membership by counting the number of different subreddits they are part of for each month from recovery start date till last recovery announcement or relapse date. Hence, richness assesses the extent to which a subject is part of multiple groups since the beginning of recovery.

Evenness represents how equally individuals are divided between types. Low values of evenness indicate the dominance of one or a few types. The way evenness is calculated is dependent on other compound diversity measures, like Shannon's diversity, such that we calculate the diversity index, and evenness is derived from that [33]. We measure the evenness by calculating the Pielou's evenness index over the frequency of subjects' posts across different subreddits [5]. Complete evenness is reached when a subject's posts are equally distributed across all subreddits in which they have posted. Here, we use evenness to assess active participation across multiple groups (rather than a focus on a single group or few groups).

We will use both richness and evenness as factors in the extended Cox model to assess the effect of multiple group memberships on relapse.

[5]Pielou's evenness is calculated as $evenness = \frac{H'}{H'_{max}}$ where H' is the Shannon diversity index, and H'_{max} is the maximum diversity which is equal to the natural logarithm of richness.

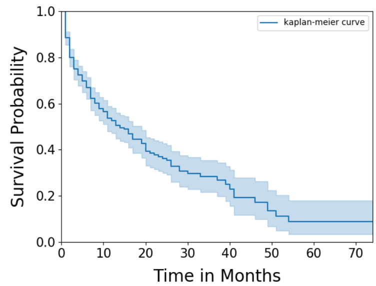

Fig. 2. Survival curve showing likelihood of experiencing a "relapse" event

6 RESULTS

6.1 Survival Analysis

The Kaplan-Meir estimator calculates the cumulative probability of not experiencing the relapse as a function of time. As we can observe in Figure 2, the sharpest decline in survival probability (probability of not experiencing a relapse) happens during the first six months after the onset of recovery. In general, probability of experiencing the relapse increases over time but at a slower rate after the first six months. Using this analysis, the median time to relapse is 12-13 months, in other words, during the first year, around 50% of the population relapse.

In order to better understand the way survival probabilities and hazard rates change over time, we also construct a life table (see Table 2). This allows for time intervals with different lengths. The 'probability density function' (pdf) of an interval is the estimate of the probability of relapsing in this interval per unit width (here: month). Hazard of an interval is an estimate of the number of relapses per month divided by the average number of survivors at the interval midpoint. Both the pdf and hazard are decreasing over time. This indicates that (1) the probability of relapsing per month is decreasing over time, and (2) the decline in average number of relapses is higher than the decline in the average number of survivors.

These two methods together provide a comprehensive explanation of how survival probability is changing over time. They are analysing it from slightly different angles, but both highlight the importance of the early months of recovery. If an addict in recovery can avoid relapse in the first six months, the chance for a sustainable recovery is much higher.

6.2 Recovery Models and Goodness of Fit

In this subsection, we report the goodness of fit of a number of different extended Cox models that estimate the survival probability of subjects in our data. With the two different measures of richness and evenness of group membership, we report three models overall. The first two models correspond to each measure separately, and the third includes both measures simultaneously. In this way, we can examine the role of each measure separately in inferring the likelihood of relapse in our data.

To evaluate the goodness of fit of the three extended Cox models we use the likelihood ratio test, or LR statistic (see Table 3). Likelihood measures how well the model is fitted to the data, with higher values indicating a better fit [6]. The LR test is performed by comparing the fit of one model to the fit of the other, in our case comparing the fit of non-null models to the null model, and to test whether the observed difference in model fit is statistically significant. In general, it can be written

[6]The log likelihood is always negative, with higher values (closer to zero) indicating a better fitting model.

Table 3. Model fit of extended Cox models, with Null as the intercept-only model.

Model	LogLikelihood	df	p-value
Null	-939.4	0	
Evenness	-935.078	1	0.003
Richness	-893.979	1	1.556e-21
full	-890.570	2	6.214e-22

Table 4. Individual Cox regression models predicting relapse from either richness or evenness of group membership.

Factor	Coef	HR=exp($Coef$)	p-value
Evenness	-1.094	0.335	0.003
Richness	-0.041	0.959	3.7e-10

in the form of LR = $2(loglik(m_2) - loglik(m_1))$ [7]. Where m_1 is the model with less parameters, the null model in our case, and m_2 is the model with added parameters, individual models for each measure and the full model. The resulting test statistic follows a chi-square distribution, with degrees of freedom equal to the number of parameters that are constrained.

As an example, comparing the log-likelihood of the full model with that of the Null model, we see that the information provided by the corresponding variables has significant explanatory power: LR = 2(-890.570 - (-939.4)), so our LR test statistic is 97.66 (distributed chi-squared), with two degree of freedom with a p-value of 6.214e-22, indicating that the full model fits significantly better than the null model.

6.3 What Predicts Relapse?

Predictors of relapse were estimated using the extended Cox model. Table 4 and 5 list the two predictors and their associated coefficients along with their hazard ratios for the individual models and the full model. The hazard ratio for a variable denotes the likelihood of a user relapsing with one unit increase in the value of the corresponding variable at any given time t. The hazard ratio formula for the extended Cox model is:

$$\widehat{HR(t)} = \frac{\hat{h}(t, X^*(t))}{\hat{h}(t, X(t))} = exp(\sum_{i=1}^{n} Coef_i(X_i^*(t) - X_i(t))) \tag{1}$$

where $X^*(t)$ and $X(t)$ identify two specifications at time t for the set of time-dependent predictors. For the full model, the hazard ratio is calculated by keeping other factors constant at time t. A hazard ratio of 1 means a lack of association, a hazard ratio smaller than 1 indicates a decreased chance of relapsing (i.e., increased survival rate), while a hazard ratio larger than 1 indicates an increased chance to relapse. We further discuss how different measures relate to the likelihood of a relapse.

Richness: Engaging with more groups during the recovery process negatively impacts the likelihood of relapse. The strongest predictor for relapse in our cohort is the number of groups a subject is engaging with. Increasing the number of online group memberships by one group, decreases the chance of relapse by 4% (Coef = -0.04, $P = 3.7e - 10$ in the individual model, and Coef

[7] $loglik$ is the natural log of the model's likelihood.

Table 5. Full Cox regression model predicting relapse from both richness and evenness of group membership

Factor	Coef	HR=exp(Coef)	p-value
Evenness	-0.90213	0.40570	0.0066
Richness	-0.04088	0.95995	2e-10

= -0.04, $P < 2e - 10$ in the full model). This result supports the prediction by SIMOR that a higher number of group memberships supports a sustained recovery.

Evenness: Actively participating in different online groups in a more even manner is negatively affecting the risk of relapse. Evenness scales between 0 and 1, therefore, a hazard ratio of 0.4057 means that increasing the evenness from its lowest to the highest value would decrease the chance of relapse by 59% ($e^{-0.90213}$ = 0.4057), and increasing the evenness by one percentage point would decrease the chance of a relapse by 1% ($e^{-0.90213*0.01}$ = 0.99). The effect of this measure is significant both individually and when richness is also included in the model. This finding supports the prediction by SIMOR that engagement (rather than purely membership) is an important factor in sustaining recovery.

Summary of Findings. We represent both individual models and the full model to see how each feature affects the survival probability as a single factor and as a pattern. By doing so, we can show whether a variable loses its importance when the effect of a related variable is controlled for. In our study, evenness and richness are significantly affecting relapse, both individually and as a pattern. This finding indicates that multiple group membership protects against relapse irrespective of the evenness of participation in each group. Similarly, evenness of participation in multiple groups protects against relapse irrespective of the absolute number of online groups an individual engages in. If a recovering addict actively participates in many subreddits, they are less likely to relapse. At the same time, engaging evenly in all these forums can additionally facilitate a successful recovery.

7 DISCUSSION

The presented study contributes to the current understanding of social aspects of addiction recovery by analysing longitudinal naturally occurring online data. Our findings point towards a positive contribution of online forums to the recovery journey of opioid addicts. In particular, our results suggest that those in recovery that reach out to a large number of different online interest groups are less likely to experience relapse than those who engage with only a small number of forum groups. In addition, a relatively even investment in these groups appears to help those in recovery from relapsing.

These findings are in line with the Social Identity Model of Recovery (SIMOR) by showing that an engagement with multiple online groups, and active participation in these groups, helps to sustain recovery from addiction [4]. Our study contributes to the growing literature on SIMOR, and the "social cure" more widely [19], by showing the important role that online forums can play. It is particularly noteworthy here that many of the online forums that subjects in our study contribute to were everyday shared interest groups rather than groups designed to facilitate addiction recovery. Non-drug related groups provide opportunities for re-defining the self outside of addiction which is thought to be a key element in creating a sustainable recovery from addiction [4].

In addition to supporting the "social cure" hypothesis that multiple group memberships improve health outcomes, our results show that active participation in these groups is an important aspect of group membership. In particular, our study measured the evenness of participation in behavioral terms by looking at the actual posting behavior of forum users. Rather than focusing on a small number of groups and establishing strong support relationships with these members, our results suggest that an even engagement with a relatively large number of groups is more effective in staving off relapse. From a social support perspective, this result appears surprising as it is the quality of interpersonal bonds to others that provide emotional and practical support during difficult times [48]. In contrast, a social identity perspective emphasises the joining of groups as a way to tip the balance away from addiction as a self-defining group membership towards alternative identities [4]. Our results appear to support the social identity perspective of investing evenly in many groups rather than concentrating on a few in the online world.

Our findings extend SIMOR by showing that online group memberships contribute positively to opioid addiction recovery, mirroring effects of offline group memberships. From a social identity perspective, this finding is not surprising, given that the Social Identity Model of Deindividuation Effects (SIDE; [26] suggests that anonymous online interactions increase a sense of group membership by reducing perceived individual differences (e.g., based on physical appearance).

7.1 Limitations

Similar to other research that follows the recovery journey of addicts rather than testing an intervention with a randomized controlled trial (RCT), our findings cannot be interpreted in a causal way. Although our findings are consistent with a leading model of recovery in the field suggesting a causal influence of group membership on maintaining recovery, we cannot conclude from our data that it is the active participation in numerous online forums itself that prevents relapse. The extent to which addicts in recovery have the resources, skills and personality to engage with numerous groups may differ between those that relapse and those that are able to maintain their recovery for longer. Factors such as interpersonal stressors, rejection sensitivity, a lack of social skills and others are known to put addicts at risk of relapse [28]. These same factors may also affect an individual's ability to engage with several online forums. Hence, engagement with a variety of online forums may simply be an indicator that the individual has the resources and skills that enable them to cope well during recovery, rather than having a direct influence on recovery itself. Nevertheless, our findings suggest that studying online forums as an effective, low-threshold intervention for addicts in recovery may be a fruitful avenue for future research.

Another limitation of our study is that it does not differentiate between social groups related to drug use and those that are not drug related. SIMOR suggests that drug-related groups may increase the risk of relapse because such groups are likely to maintain a sense of self as an addict in the recovering user, thereby acting as a "social curse" rather than a "social cure" [14]. In contrast, groups supporting recovery may instill a sense of self as a recovering addict, providing more supportive norms and a more positive view of the self. A study by [8] shows that thinking of oneself as a recovering, or recovered, addict is negatively associated with self-reported relapse. Similarly, social groups unrelated to drug use are seen by SIMOR as a key resource to allow addicts to move away from addiction in the long term. Further analysis is needed that disentangles which online groups and forums are helpful in maintaining recovery, and which ones have the potential to be harmful.

A third limitation of our method is that it cannot take into account group membership outside of Reddit forums. Subjects in our study are highly likely to have social connections beyond Reddit forums, such as participating in other online groups (e.g., on social media) and in various face-to-face groups. Interestingly, a recent self-report study that used social identity mapping to understand the extent of the social networks of addicts indicates a much smaller number of social groups (Med = 4) than our method of counting contributions to different Reddit subforums (Med = 14) [2]. One reason for the discrepancy may be that online forums are not reported as 'social contacts' by participants [23] - perhaps due to the lack of face-to-face interaction or perhaps because of their relatively mundane nature (e.g., exchanging views on crime novels or sharing BBQ recipes).

Importantly, the social identity approach favours a wide definition of social identity where a group or category membership consists of two or more individuals who perceive the grouping to be of value and emotional significance [44]. Another reason for the discrepancy with self-reported mappings may be differences in the participant sample, such that our sample is clearly self-selected towards online engagement whereas samples recruited through treatment groups may be less active socially, at least online. For those active in online forums, our findings suggest that investing in numerous online groups is beneficial for maintaining recovery from opiate addiction.

7.2 Implications

7.2.1 Computational Social Science Contribution. Our findings show that computational social science methods, and the use of online data, can inform the study of addiction recovery with information difficult to access through traditional clinical or psychological studies. Using survival analysis, we can identify prospective factors associated with the likelihood of relapse in a large sample; thus we go beyond retrospective analyses of risk factors obtained from self-identified clinical in-patients. Finally, we can longitudinally forecast changes in whether and when subjects will begin to relapse.

7.2.2 Practical Implications. The effectiveness of easy-to-join online groups is of particular importance for addicts who have no, or limited, access to face-to-face groups. Such limited access may be due to physical barriers such as living in remote areas, being physically disabled, having work and/or caring commitments, or needing to self-isolate for health reasons (e.g., during the Covid-19 pandemic). Similarly, psychological barriers such as worries about being stigmatized, social anxiety, or depression can make it difficult to join and commit to face-to-face groups on a regular basis. Being active in a variety of online forums may provide an additional resource to support those in recovery, potentially making 'online social prescribing' a viable addition, or in some cases an alternative, to joining face-to-face groups [24].

To provide opportunities for 'online social prescribing', online platforms that are currently solely providing a peer-support forum discussing recovery may wish to consider broadening their offer to allow sub-groups on a variety of topics to form. This would allow those that do not feel confident in joining online communities on 'mainstream' platforms the opportunity to acquire a range of online community memberships. Having several communities on the same platform would also facilitate an even engagement as it reduces the additional step for platform users of having to monitor conversations on several platforms.

7.3 Future Directions

In order to test whether, and for whom, 'online social prescribing' is a viable alternative to face-to-face groups, future research is needed that can assess the full picture of online and offline group activities of recovering addicts. However, such a complete tracing of activities is extremely difficult to study as self-reporting (e.g., in a diary study) may still underestimate the number of different online - and potentially also offline - groups an individual contributes to. Similarly, naturally occurring data is limited as anonymity and privacy preclude us from tracking subjects across different platforms.

Beyond the requirement for large-scale assessments of 'online social prescribing', future research may first wish to understand how different types of groups (e.g., addiction/drug-related groups, recovery support groups, and non-drug-related groups) relate to relapse. Even though many groups can serve as a 'social cure', there are also groups that may act as a 'social curse' by reinforcing behaviours related to drug use and/or by affecting mental health negatively.

Another area for future research relates to the assertion by SIMOR that group membership changes the self-perception of recovering addicts over time, and that this is the key mechanism driving sustainable recovery. Traditionally, social groups have been seen as providing practical advice and emotional support, thereby enabling addicts to progress through their recovery journey. Although SIMOR recognises the utility of social networks, it suggests that the fundamental change that needs to take place is for the recovering addict to think about themselves less in terms of being an addict, and more in terms of being 'recovered' as well as acquiring other non-drug related identities over time. Currently, there is no validated method to study such changes in identity salience for naturally occurring data.

8 CONCLUSION

Using survival analysis on naturally occurring data, this study provides initial support that multiple online group memberships, and active participation in several online groups, can reduce the risk of relapse in recovering opioid addicts significantly. It supports a basic assertion of the social identity model of recovery (SIMOR) that social groups are a key element in sustaining recovery from addiction. The finding also points towards 'online social prescribing' as a potential alternative to joining face-to-face groups.

9 ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council UK through a fellowship to the second author [EP/S001409/1]. The authors would like to thank Rada Biyukova and Jodie Hall for their help with data labeling. Parts of this manuscript were presented at the IC2S2 conference 2020 [8].

REFERENCES

[1] Robert D Ashford, Salvatore Giorgi, Beau Mann, Chris Pesce, Lon Sherritt, Lyle Ungar, and Brenda Curtis. 2020. Digital recovery networks: Characterizing user participation, engagement, and outcomes of a novel recovery social network smartphone application. *Journal of Substance Abuse Treatment* 109 (2020), 50–55. https://doi.org/10.1016/j.jsat.2019.11.005

[2] Melinda Beckwith, David Best, Michael Savic, Catherine Haslam, Ramez Bathish, Genevieve Dingle, Jock Mackenzie, Petra K. Staiger, and Dan I. Lubman. 2017. Social Identity Mapping in Addiction Recovery (SIM-AR): Extension and application of a visual method. *Addiction Research & Theory* 27, 6 (2017), 462–471. https://doi.org/10.1080/16066359.2018.1544623

[3] Brandon G Bergman, Nathaniel W Kelly, Bettina B Hoeppner, Corrie L Vilsaint, and John F Kelly. 2017. Digital recovery management: Characterizing recovery-specific social network site participation and perceived benefit. *Psychology of Addictive Behaviors* 31, 4 (2017), 506. https://doi.org/10.1037/adb0000255

[4] David Best, Melinda Beckwith, Catherine Haslam, S Alexander Haslam, Jolanda Jetten, Emily Mawson, and Dan I Lubman. 2016. Overcoming alcohol and other drug addiction as a process of social identity transition: The social identity model of recovery (SIMOR). *Addiction Research & Theory* 24, 2 (Aug. 2016), 111–123. https://doi.org/10.3109/16066359.2015.1075980

[5] David Best, Andrew Rome, Kirstie A Hanning, William White, Michael Gossop, Avril Taylor, and Andy Perkins. 2010. *Research for recovery: A review of the drugs evidence base.* Crown, Edinburgh, United Kingdom. https://www2.gov.scot/resource/doc/321958/0103435.pdf

[6] Ana-Maria Bliuc, David Best, Muhammad Iqbal, and Katie Upton. 2017. Building addiction recovery capital through online participation in a recovery community. *Social Science & Medicine* 193 (2017), 110–117. https://doi.org/10.1016/j.socscimed.2017.09.050

[7] Ana-Maria Bliuc, Tuyet-Ngan Doan, and David Best. 2019. Sober social networks: The role of online support groups in recovery from alcohol addiction. *Journal of Community & Applied Social Psychology* 29, 2 (Nov 2019), 121–132. https://doi.org/10.1002/casp.2388

[8] http://2020.ic2s2.org/6th-international-conference-computational-social-science

[8] Sarah A Buckingham, Daniel Frings, and Ian P Albery. 2013. Group membership and social identity in addiction recovery. *Psychology of Addictive Behaviors* 27, 4 (2013), 1132. https://doi.org/10.1037/a0032480

[9] Centre for Disease Control and Prevention. 2018. About underlying cause of death 1999-2018. https://wonder.cdc.gov/ucd-icd10.html

[10] Stevie Chancellor and Munmun De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: A critical review. *npj Digital Medicine* 3, 43 (Mar 2020). https://doi.org/10.1038/s41746-020-0233-7

[11] E. Clark R., J. D. Baxter, G. Aweh, E. O'Connell, W. H. Fisher, and B. A. Barton. 2015. Risk factors for relapse and higher costs among Medicaid members with opioid dependence or abuse: Opioid agonists, comorbidities, and treatment history. *Journal of Substance Abuse Treatment* 57 (Oct 2015), 59–80. https://doi.org/10.1016/j.jsat.2015.05.001

[12] Glen Coppersmith, Kim Ngo, Ryan Leary, and Anthony Wood. 2016. Exploratory analysis of social media prior to a suicide attempt. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*. 106–117.

[13] Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3267–3276.

[14] Genevieve A. Dingle, Claire Stark, Tegan Cruwys, and David Best. 2015. Breaking good: Breaking ties with social groups may be good for recovery from substance misuse. *British Journal of Social Psychology* 54, 2 (2015), 236–254. https://doi.org/10.1111/bjso.12081

[15] Chad D Emrick, J Scott Tonigan, Henry Montgomery, and Laura Little. 1993. Alcoholics Anonymous: What is currently known? In *Research on Alcoholics Anonymous: Opportunities and alternatives*, W. R. Miller B. S. McCrady (Ed.). Rutgers Center of Alcohol Studies, 41–76.

[16] Seyedezahra Shadi Erfani, Yvette Blount, and Babak Abedin. 2016. The influence of health-specific social network site use on the psychological well-being of cancer-affected people. *Journal of the American Medical Informatics Association* 23, 3 (Jan 2016), 467–476. https://doi.org/10.1093/jamia/ocv170

[17] Ryan Eshleman, Deeptanshu Jha, and Rahul Singh. 2017. Identifying individuals amenable to drug recovery interventions through computational analysis of addiction content in social media. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Kansas City, MO, USA). IEEE, 849–854. https://doi.org/10.1109/BIBM.2017.8217766

[18] Simon Graham, Jamie Irving, Ivan Cano, and Michael Edwards. 2018. Participation with online recovery specific groups-findings from the UK life in recovery survey 2015. *Alcoholism Treatment Quarterly* 36, 4 (2018), 459–481.

[19] Catherine Haslam, Jolanda Jetten, Tegan Cruwys, Genevieve A. Dingle, and S. Alexander Haslam. 2018. *The New Psychology of Health: Unlocking the Social Cure.* Routledge, Abingdon, United Kingdom.

[20] Mette Terp Høybye, Christoffer Johansen, and Tine Tjørnhøj-Thomsen. 2005. Online interaction: Effects of storytelling in an internet breast cancer support group. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer* 14, 3 (Jul 2005), 211–220. https://doi.org/10.1002/pon.837

[21] Kyle Kampman and Margaret Jarvis. 2015. American Society of Addiction Medicine (ASAM) National Practice Guideline for the Use of Medications in the Treatment of Addiction Involving Opioid Use. *Journal of Addiction Medicine* 9, 5 (2015), 358–367. https://doi.org/10.1097/ADM.0000000000000166

[22] Edward L Kaplan and Paul Meier. 1958. Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* 53, 282 (1958), 457–481. https://doi.org/10.1080/01621459.1958.10501452

[23] Tina F. Keil and Miriam Koschate. in press. Variations in subjective definitions of everyday situations as intergroup contact. *British Journal of Social Psychology* (in press). https://doi.org/10.1111/bjso.12372

[24] Richard Kimberlee. 2016. What is the value of social prescribing? *Advances in Social Sciences Research Journal* 3, 3 (2016). https://doi.org/10.14738/assrj.33.1889

[25] Wendy Kissin, Caroline McLeod, and James McKay. 2003. The longitudinal relationship between self-help group attendance and course of recovery. *Evaluation and Program Planning* 26, 3 (Aug 2003), 311–323. https://doi.org/10.1016/S0149-7189(03)00035-1

[26] Olivier Klein, Russell Spears, and Stephen Reicher. 2007. Social Identity Performance: Extending the Strategic Side of SIDE. *Personality and Social Psychology Review* 11, 1 (2007), 28–45. https://doi.org/10.1177/1088868306294588

[27] David G Kleinbaum and Mitchel Klein. 2012. Extension of the Cox proportional hazards model for time-dependent variables. In *Survival analysis*. Springer, Chapter 6, 241–288. https://link.springer.com/chapter/10.1007/0-387-29150-4_6

[28] David Leach and Henry R Kranzler. 2013. An interpersonal model of addiction relapse. *Addictive disorders & their Treatment* 12, 4 (2013), 183. https://doi.org/10.1097/ADT.0b013e31826ac408

[29] John Lu, Sumati Sridhar, Ritika Pandey, Mohammad Al Hasan, and Georege Mohler. 2019. Investigate Transitions into Drug Addiction through Text Mining of Reddit Data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Anchorage AK USA). ACM, 2367–2375. https://doi.org/10.1145/3292500.3330737

[30] Diana MacLean, Sonal Gupta, Anna Lembke, Christopher Manning, and Jeffrey Heer. 2015. Forum77: An analysis of an online health forum dedicated to addiction recovery. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver BC Canada). ACM, 1511–1526. https://doi.org/10.1145/2675133.2675146

[31] Anne E Magurran. 2013. *Measuring biological diversity*. John Wiley & Sons.
[32] Rudolf H Moos and Bernice S Moos. 2006. Participation in treatment and Alcoholics Anonymous: A 16-year follow-up of initially untreated individuals. *Journal of Clinical Psychology* 62, 6 (Mar 2006), 735–750. https://doi.org/10.1002/jclp.20259
[33] E Kathryn Morris, Tancredi Caruso, François Buscot, Markus Fischer, Christine Hancock, Tanja S Maier, Torsten Meiners, Caroline Müller, Elisabeth Obermaier, Daniel Prati, et al. 2014. Choosing and using diversity indices: Insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution* 4, 18 (Aug 2014), 3514–3524. https://doi.org/10.1002/ece3.1155
[34] Elizabeth L Murnane and Scott Counts. 2014. Unraveling abstinence and relapse: Smoking cessation reflected in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto Ontario Canada). ACM, 1345–1354. https://doi.org/10.1145/2556288.2557145
[35] E. V. Nunes, M. Gordon, P. D. Friedmann, M. Fishman, J. D. Lee, D. T. Chen, M.C. Hu, T. Y. Boney, D. Wilson, and C. P. O'Brien. 2018. Relapse to opioid use disorder after inpatient treatment: Protective effect of injection naltrexone. *Journal of Substance Abuse Treatment* 85 (Feb 2018), 49–55. https://doi.org/10.1016/j.jsat.2017.04.016
[36] Michael J Paul and Mark Dredze. 2017. Social monitoring for public health. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 9, 5 (2017), 1–183.
[37] Priya Ranganathan and C. S. Pramesh. 2012. Censoring in survival analysis: potential for bias. *Perspectives in Clinical Research* 3, 1 (Mar 2012), 40. https://doi.org/10.4103/2229-3485.92307
[38] Karen Rodham, Candy McCabe, and David Blake. 2009. Seeking support: An interpretative phenomenological analysis of an internet message board for people with Complex Regional Pain Syndrome. *Psychology and Health* 24, 6 (Nov 2009), 619–634. https://doi.org/10.1080/08870440802563245
[39] Fabio Sani. 2012. Group identification, social relationships, and health. In *The social cure: Identity, health and well-being*, Jolanda Jetten, Catherine Haslam, and Alexander S. Haslam (Eds.). Psychology Press New York, NY, Chapter 2, 21–37.
[40] Michael Savic, David Best, Simone Rodda, and Dan I Lubman. 2013. Exploring the focus and experiences of smartphone applications for addiction recovery. *Journal of Addictive Diseases* 32, 3 (Jul 2013), 310–319. https://doi.org/10.1080/10550887.2013.824331
[41] Patrick Schober and Thomas R Vetter. 2018. Survival analysis and interpretation of time-to-event data: The tortoise and the hare. *Anesthesia and Analgesia* 127, 3 (July 2018), 792. https://doi.org/10.1213/ANE.0000000000003653
[42] Julia MA Sinclair, Sophia E. Chambers, and Christopher C. Manson. 2017. Internet support for dealing with problematic alcohol use: a survey of the Soberistas online community. *Alcohol and alcoholism* 52, 2 (2017), 220–226.
[43] Meredith M Skeels, Kenton T Unruh, Christopher Powell, and Wanda Pratt. 2010. Catalyzing social support for breast cancer patients. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta Georgia USA). ACM, 173–182. https://doi.org/10.1145/1753326.1753353
[44] Henri Tajfel. 1981. *Human groups and social categories: Studies in social psychology*. Cup Archive.
[45] Acar Tamersoy, Duen Horng Chau, and Munmun De Choudhury. 2017. Analysis of smoking and drinking relapse in an Online community. In *Proceedings of the 2017 International Conference on Digital Health*. 33–42.
[46] Acar Tamersoy, Munmun De Choudhury, and Duen Horng Chau. 2015. Characterizing smoking and drinking abstinence from social media. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*. 139–148.
[47] U. S. Department of Health and Human Services. 2017. HHS Acting Secretary Declares Public Health Emergency to Address National Opioid Crisis. https://www.hhs.gov/about/news/2017/10/26/hhs-acting-secretary-declares-public-health-emergency-address-national-opioid-crisis.html
[48] Debra Vandervoort. 1999. Quality of social support in mental and physical health. *Current Psychology* (1999), 205–221. https://doi.org/10.1007/s12144-999-1029-8
[49] Nikhita Vedula and Srinivasan Parthasarathy. 2017. Emotional and linguistic cues of depression from social media. In *Proceedings of the 2017 International Conference on Digital Health*. 127–136.
[50] Nora D. Volkow. 2020. Collision of the COVID-19 and Addiction Epidemics. *Annals of Internal Medicine* (Apr 2020). https://doi.org/10.7326/M20-1212
[51] Robert H Whittaker. 1972. Evolution and measurement of species diversity. *Taxon* 21, 2-3 (May 1972), 213–251. https://doi.org/10.2307/1218190
[52] World Health Organisation. 2018. Information sheet on opioid overdose. https://www.who.int/substance_abuse/information-sheet/en/
[53] World Health Organization. 2001. *The World Health Report 2001 - Mental health: New understanding, new hope*. World Health Organization. https://www.who.int/whr/2001/en/
[54] Yue Zhang, Arti Ramesh, Jennifer Golbeck, Dhanya Sridhar, and Lise Getoor. 2018. A structured approach to understanding recovery and relapse in AA. In *Proceedings of the 2018 World Wide Web Conference*. 1205–1214.

Received June 2020; revised October 2020; accepted December 2020