On the Difficulty of Predicting Engagement with Digital Health for Substance Use

Franziska GÜNTHER[a,1], Christopher YAU[b,c] , Sarah ELISON-DAVIES[d]
and David WONG[a]

[a] *Centre for Health Informatics, School of Health Sciences, University of Manchester, UK*

[b] *Nuffield Department of Women's and Reproductive Health, University of Oxford, UK*

[c] *Health Data Research, UK*

[d] *LifeWorks, Canada*

ORCiD ID: Franziska Günther https://orcid.org/0000-0002-6016-765X

Abstract. Digital interventions can be an important instrument in treating substance use disorder. However, most digital mental health interventions suffer from early, frequent user dropout. Early prediction of engagement would allow identification of individuals whose engagement with digital interventions may be too limited to support behaviour change, and subsequently offering them support. To investigate this, we used machine learning models to predict different metrics of real-world engagement with a digital cognitive behavioural therapy intervention widely available in UK addiction services. Our predictor set consisted of baseline data from routinely-collected standardised psychometric measures. Areas under the ROC curve, and correlations between predicted and observed values indicated that baseline data do not contain sufficient information about individual patterns of engagement.

Keywords. prediction, digital health, substance use, engagement

1. Introduction

Digital interventions (DIs) for people with substance use disorders (SUDs) are digitalised complements or temporary replacements of traditional face-to-face therapies such as cognitive behavioural therapy (CBT). With DIs being more cost-effective and 24/7 accessible, they can represent an important instrument in treating SUDs.

To derive improved mental health outcomes via a DI, users need to engage with DI content to a sufficient degree. However, maintaining user engagement has been a consistent problem for DIs for mental health [1]. Early, accurate prediction of level of DI engagement could allow users at high risk of poor engagement to be identified. This could potentially be used to target additional support. The basis of such prediction should be data collected early into the user journey, if feasible at first user contact with a DI, as dropout after first use is a common phenomenon. However, it is not clear if prediction is at all possible using such data, since real-world engagement may depend on multiple factors that may not be reflected in a one-off clinical assessment before engagement.

The aim of this study is hence to assess whether engagement with one DI called Breaking Free Online (BFO) can be predicted using data collected at first DI use.

---

[1] Corresponding Author: Franziska Günther, E-Mail: franziska.gunther@postgrad.manchester.ac.uk.

2. Methods

2.1. Source of data

Data were routinely collected from BFO users enrolled between July 2016 and October 2022 in 513 community addiction services (CAS) in the UK. BFO is a self-guided digital CBT programme for SUDs, which for the past decade has been widely available to clients of CAS in the UK. Ethical approval for collection, storage and use of data accumulating from routine use of BFO by clients in participating treatment services, was obtained from an NHS Research Ethics Committee (London - South East, 16 May 2012 and 22 May 2017, references 12/LO/0076 and 12/LO/0287).

The BFO programme features six modules, with each split into one part psychoeducation and one part practice (applying what was learned in psychoeducation to one's own life). These subparts are subsequently referred to as "strategies", specifically, *information strategies* and *action strategies*.

Users are required to complete a baseline assessment so that modules can later be recommended to them. The baseline assessment includes four questionnaires: the Severity of Dependence Scale (SDS) [2], the Patient Health Questionnaire 4 [3], the World Health Organization Quality of Life measure (items 1, 2, 17, 18, and 20) [4] and the Recovery Progression Measure [5]. In addition, the baseline assessment also recorded user age, gender, ethnicity, abused substances, substance-using days in the preceding week and the user's target for substance-free days per week.

In addition to the assessment questionnaire data, assessment dates as well as module completion data, specifically the number of *information* and *action strategy* completions for each programme module and their most recent completion dates, were also available.

2.2. Predictors

The feature set we used to develop a prediction model of BFO engagement comprises 62 features corresponding to every question in the baseline assessment and a set of derived variables, specifically (1) baseline abstinence defined as zero substance-using days per week, (2) the number of days from registration to first assessment completion, (3) the number of clinical complexity inducing factors present at baseline (including financial difficulties, cravings, difficulties with physical health, at work, or with housing) and (4) cutoff-based variables on baseline anxiety, depression and substance dependence.

2.3. Outcomes

We used 9 continuous variables as outcomes with each measuring a different aspect of user engagement, as follows: (1) the number of days from the first to the last use event, subsequently referred to as the number of accessed days, (2) the number of strategies completed, (3) the number of *information strategies* completed, (4) the number of *action strategies* completed, (5) the number of use events (all assessments + strategies completed), (6) the use rate (number of use events / number of accessed days), (7) the percentage of days actively engaged (with the number of days on which an assessment was completed - which empirically fall together with known days of module completion in 67% percent of cases - regarded as active engagement), (8) the median intermission length in days (with days on which no active engagement was registered described as intermission days) and (9) the mean absolute deviation (MAD) intermission length. Log-

transformation was applied to all these variables. We also used the completion of 8 or more strategies as a binary outcome, referring to 8 sessions as the dose of talking therapy commonly received by patients completing a course of treatment through the NHS [6].

2.4. Statistical analysis and missing data

We predicted the 9 continuous outcomes and 1 binary outcome independently, using random forests and the XGBoost algorithm out-of-the-box with 10-fold cross validation. Stratification was applied to the target variable, with numeric strata being binned into quartiles. The average area under the receiver operating curve was used as a measure of predictive performance for binary outcome variables. Correlations between the observed and predicted values served as an assessment of predictive performance for the continuous outcome variables. The average root mean squared error (RMSE) was used to compare predictive performance between random forests and the XGBoost algorithm. We removed data from users who had >80% data missing on their baseline assessment (n = 706) as multiple imputation would be difficult for these users. For the remainder of the data, we opted for a complete case analysis as only 5% of these cases had incomplete data, and < 4% of cells were missing in total. R code for this analysis is available at https://github.com/franziskagunther/predict-engagement.

Table 1. User characteristics at baseline.

Characteristic	Statistic/Label	Value
Age in years	mean (SD, range)	40.1 (11.7, 18 - 84)
Gender	Female	47.1% (10,745)
	Male	52.5% (11,967)
	Other	0.3% (79)
Ethnicity	White	93% (21,207)
	Asian / Asian British	1.9% (426)
	Black / Black British	1.7% (382)
	Mixed	2.7% (626)
	Other	0.7% (150)
Primary substances	Alcohol	63.8% (14,533)
	Cocaine	11.7% (2,659)
	Marijuana	7.9% (1,810)
	Heroin	5.5% (1,248)
	Crack	3.5% (805)
	Other (46 other substances)	7.6% (1,741)
Substance dependence (SDS sum score equal to or larger than 3)	Yes	92.6% (20,494)
	No	7.4% (1,631)
Anxiety (sum of first two PHQ-4 items equal to or larger than 3)	Yes	69% (15,593)
	No	31% (7,007)
Depression (sum of last two PHQ-4 items equal to or larger than 3)	Yes	66.5% (15,023)
	No	33.5% (7,577)
Substance-using days in the past week	modes	0: 24.5%, 7: 38.1%

3. Results

We removed users who were younger than 18 (n = 82) or older than 89 years (n = 3). We also excluded users reporting a goal of increasing their substance consumption (n = 1,314, possibly due to user interpretation of item as the desired number of substance-consuming instead of substance-free days). Finally, we excluded users whose reports of

daily substance consumption was deemed to be clinically infeasible (n = 92). The final dataset included 22,796 users. Table 1 summarises their baseline characteristics.

We first examined individual feature-outcome correlations and found these to be low (see Figure 1). Cross-feature correlations were high, instead. Finally, we used XGBoost and random forests to see if the combination of predictors could predict outcomes, but predictive performance was poor in all cases. We report prediction accuracy with the random forests which performed slightly better than XGBoost with regards to RMSE and AUC. We obtained an average AUC of 0.57 [CI: 0.56-0.58] for the prediction of completing $n>=8$ modules. Model performance did not improve for other values of n. Predictive performance for continuous outcomes was similarly low and correlations between observed and predicted outcomes ranged between 0.03 and 0.13.

4. Discussion

Prediction of real-world engagement in self-guided DIs for mental health could contribute to overcoming one of the field's biggest problems; early and frequent dropout. Many DIs routinely administer assessments on users' clinical characteristics before providing access to DI content which, in theory, represent easily obtained sets of predictors of possibly non-beneficial engagement at the earliest possible time point.

We conducted a prediction study with data from the BFO programme in which all users, regardless of their actual level of engagement with the system, were included in the analysis. State-of-the-art prediction models were unable to accurately predict a range of engagement metrics from baseline assessment data which represents evidence against the predictability of engagement with BFO from clinical information at first access.

Multiple unmeasured factors may make prediction of user engagement challenging, such as the clinical complexity of individuals with SUD which likely interferes with engagement. Given the lack of prediction accuracy in this study, triaging new CAS clients for BFO use on the basis of their baseline assessment data may exclude individuals who may engage and possibly benefit from BFO if introduced to it.

This research has some limitations. First, our metrics of engagement were behavioural, and do neither reflect cognitive or emotional involvement of users with the BFO programme nor benefit which may be achieved after minimal engagement. However, by including continuous engagement variables, we attempted to reflect that beneficial engagement can have individually different outlooks which is often ignored in studies using "minimal engagement" thresholds. Our decision of allowing for a variety of different engagement patterns also resulted in intentionally not removing engagement outliers, which may bias our outcomes.

This study focused on a single DI. Examination of other DIs is desirable but complicated by access to commercially-sensitive data for independent researchers.

5. Conclusion

Early prediction of engagement is desirable in digital mental health. Our case study of prediction modeling of engagement in digital CBT for SUD suggests that information beyond clinical baseline characteristics is necessary to achieve accurate predictions.
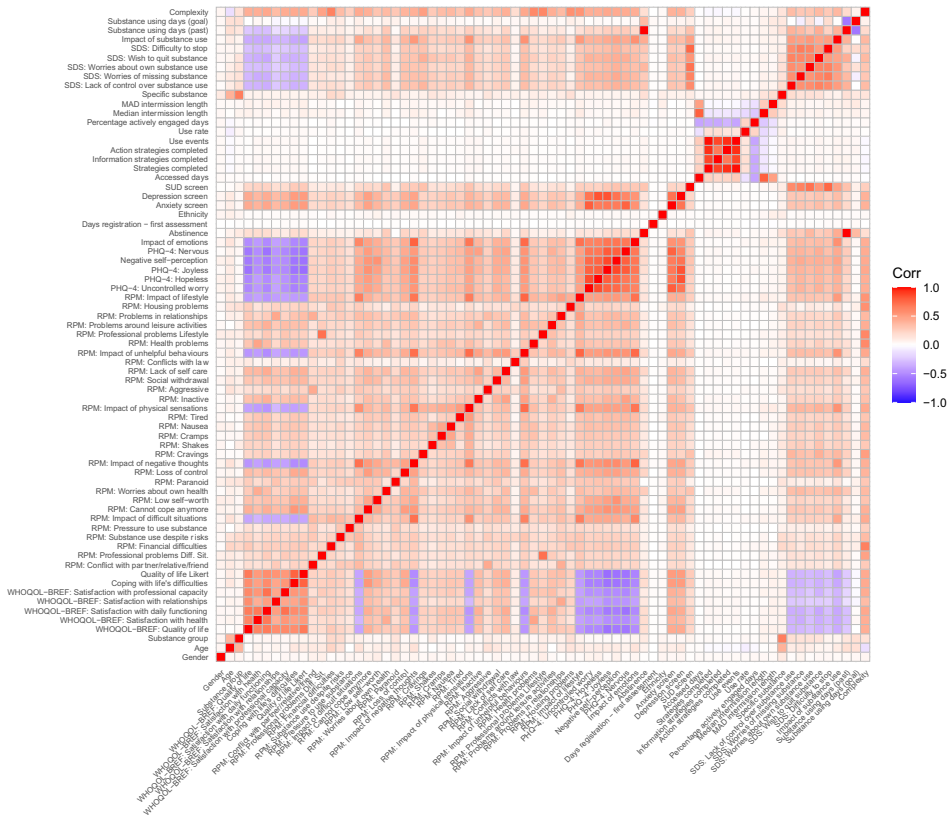
Figure 1. Correlations between variables used for prediction modelling.

References

[1] Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health. 2018 Aug;21(3):116-9, doi: 10.1136/eb-2018-102891

[2] Gossop M, Darke S, Griffiths P et al. The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. Addiction. 1995 May;90(5):607-14, doi: 10.1046/j.1360-0443.1995.9056072.x

[3] Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ–4. Psychosomatics. 2009 Nov-Dec;50(6):613-21, doi: 10.1176/appi.psy.50.6.613

[4] Skevington SM, Lotfy M, O'Connell KA. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL Group. Quality of Life Research. 2004 Mar;13(2):299-310, doi: 10.1023/B:QURE.0000018486.91360.00

[5] Elison S, Davies G, Ward J. Initial development and psychometric properties of a new measure of substance use disorder "recovery progression": the Recovery Progression Measure (RPM). Substance Use & Misuse. 2016 Jul;51(9):1195-206, doi: 10.3109/10826084.2016.1161052

[6] Population Health, Clinical Audit and Specialist Care Team, NHS Digital. Psychological therapies, annual report on the use of IAPT services, 2021-22 [Internet]. Leeds: NHS Digital; 2022 Sep 29. Report No. 12: Available from: https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2021-22