On the difficulty of predicting engagement with digital interventions for substance use disorders

Franziska GÜNTHERa,1 , Christopher YAUb,c , Sarah ELISON-DAVIESd and David C WONGa

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

b *Nuffield Department of Women’s and Reproductive Health, University of Oxford*

c *Health Data Research UK*

d *LifeWorks, Canada*

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

**Abstract.** Self-guided digital interventions can be an important instrument in treating substance use disorder. However, most digital mental health interventions suffer from early and 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 offer them greater support. To investigate this, we used standard 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 set of predictors consisted of baseline data from routinely-collected standardised psychometric measures, and variables derived from them. Areas under the ROC curve, and correlations between predicted and observed values indicated that baseline data does 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 equivalents of traditional face-to-face therapies such as cognitive behavioural therapy (CBT). They are used to complement or temporarily replace equivalent in-person interventions. 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 [1]. However, maintaining user engagement has been a consistent problem for DIs for mental health [2]. Early and 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. Prediction of poor engagement on the basis of data collected early into the user journey, if feasible at first user contact with a DI, would make targeted additional support especially effective 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 [3] that may not be reflected in a one-off clinical assessment before user engagement. However, engagement is a prerequisite for beneficial user outcomes, and hence an important prediction target, maybe even more so than clinical outcomes of DI use which have been targeted with varying success in previous prediction studies [4].

The aim of this study is therefore to assess whether engagement with one DI called Breaking Free Online (BFO) can be predicted using data routinely collected at users’ first interaction with the programme.

# 2. Methods

## 2.1. Source of data

Data were routinely collected from users of BFO enrolled between July 2016 and October 2022. Enrolment took place in 513 different community-based addiction services in the UK. BFO is a modular, self-guided digital CBT programme for SUDs, which for the past decade has been widely available to clients of community addiction services 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, 22 May 2017, reference 12/LO/0287).

The BFO programme features six modules; each module is split into one part *psychoeducation* and one complementary 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 validated questionnaires designed to measure different dimensions of SUDs: (i) the Severity of Dependence Scale [5], (ii) the Patient Health Questionnaire 4 [6], (iii) the World Health Organization Quality of Life measure (items 1, 2, 17, 18, and 20) [7] and (iv) the Recovery Progression Measure [8]. Responses to questions were recorded on 2-, 4-, 5- and 11-point Likert scales, respectively. 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.

Users struggling with both alcohol and drug dependence are required to answer questions on their alcohol and drug dependence separately. For our study we only used answers to questions on the substance labeled as the primary dependence. If users had not indicated which substance was their primary one, primacy was determined through the highest total SDS score, or if these were equal, through a greater number of substance-using days per week. If these were again equal for both substances, primacy was chosen at random.

In addition to the assessment questionnaire data, dates of user assessments as well as module completion data, specifically the number of completions for the *psychoeducation* and *practice* part of each programme module and the date of its most recent completion, were also available.

## 2.2. Predictors

The feature set we used for the development of a prediction model of BFO engagement comprises all 62 items corresponding to every question in the baseline assessment and a set of derived variables. We derived the following variables: (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 for a user (counting in the presence of financial difficulties, cravings, difficulties with physical health, at work, or with housing) and (4) cutoff-based variables on anxiety, depression and substance dependence.

## 2.3. Outcomes

We used 9 derived variables as continuous outcome measures. Each outcome measured 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 continuous outcomes. In addition, we used the completion of 8or more strategies as a binary outcome variable. We used this threshold because 8 sessions was the dose of talking therapy commonly received by patients completing a course of treatment through the NHS in England [9].

## 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 with 10-fold cross validation. Stratification was applied to the target variable, with numeric strata being binned into quartiles. Both algorithms were used out-of-the-box without hyperparameter tuning. In the case of XGBoost, all discrete features were one-hot-encoded. The average area under the receiver operating curve was used as a measure of predictive performance for binary engagement outcome variables. Correlations between the observed and predicted values served as an assessment of predictive performance for the continuous engagement 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 (except for intermission length related outcomes which were only available for those updating their assessment at least once), and < 4% of cells were missing in total. All analyses were conducted using R (version 4.2.1), and code is available at https://github.com/franziskagunther/predict-engagement.

# 3. Results

We removed users who were younger than 18 (n = 82) or older than 89 years (n = 3). We also excluded users reporting alcohol consumption of more than 100 standard units of alcohol on a typical day (n = 88), and those reporting a goal of increasing their substance consumption (n = 1314, possibly due to erroneous user interpretation of item as the desired number of substance-consuming instead of substance-free days). Finally, we excluded users whose reports of daily drug consumption was deemed to be clinically infeasible (n = 4). The final dataset contained data from 22,796 users. Table 1 summarises their baseline characteristics.

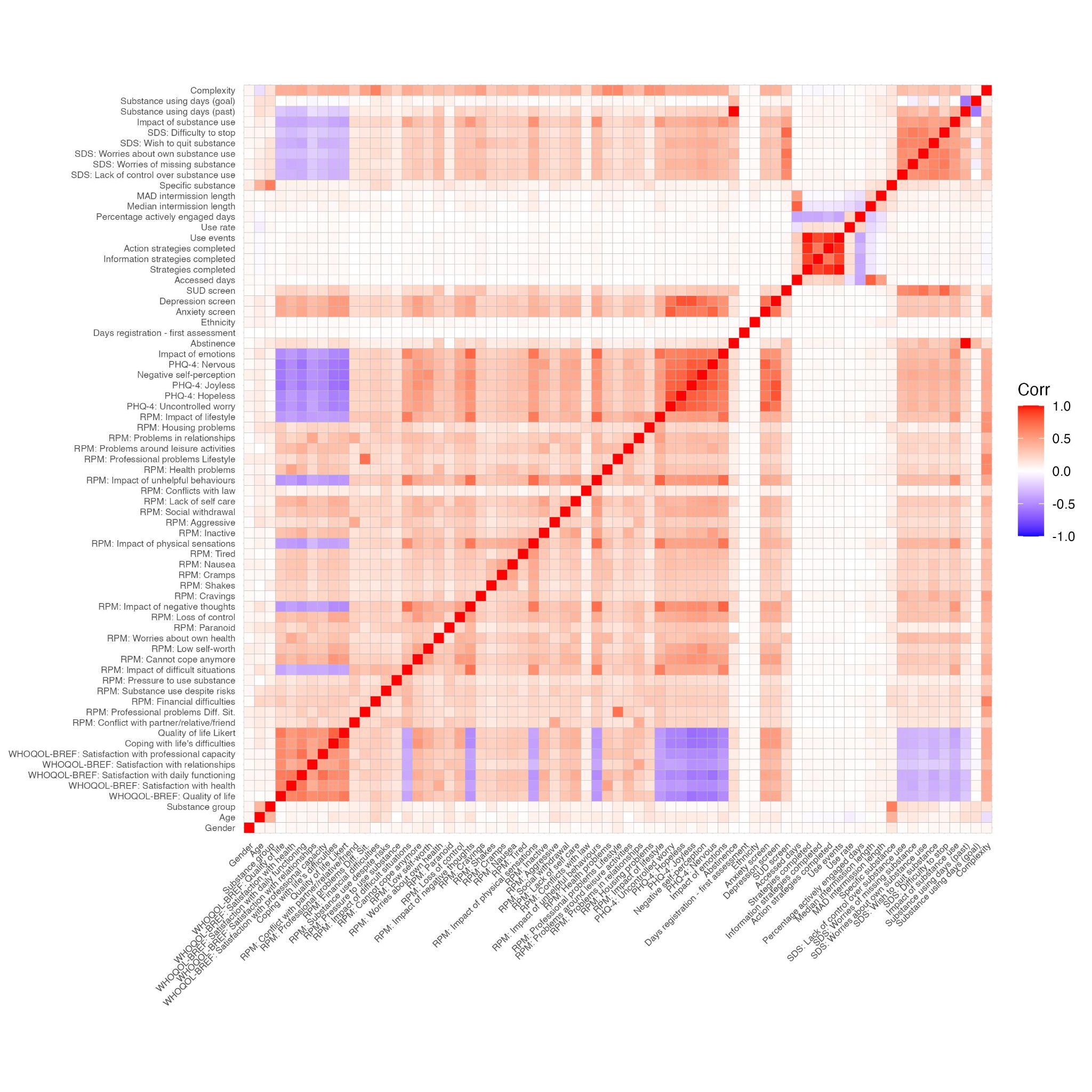
**Table 1.** User characteristics at baseline.

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

We first examined individual feature-outcome correlations and found these to be low (see Figure 1, Pearson correlation used for continuous variables, Cramer’s V for nominal variables, polychoric correlation for ordinal variables, and square root of R2 for continuous - discrete variable combinations). 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.

Because the random forest showed slightly better out-of-the-box performance than the XGBoost algorithm across outcomes with regards to RMSE and AUC, we report results for it here. We obtained an average AUC of 0.57 [CI: 0.56-0.58] for the prediction of completing *n=8* or more modules. Model performance did not improve when other values of *n* were tested (AUC for the prediction of completing one or more modules: 0.54 [CI: 0.54-0.55]).

Predictive performance for continuous outcomes was similarly low and correlations between observed and predicted outcomes range between 0.03 and 0.13.



**Figure 1.** Correlations between variables used for prediction modelling.

# 4. Discussion

Prediction of engagement in self-guided DIs for mental health, including those for the treatment of SUD, 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, in order to recommend content and to establish a baseline for evaluation of symptom change. Such assessments, in theory, represent routinely obtained sets of predictors of possibly non-beneficial engagement at the earliest possible time point.

We conducted a prediction study with real-world data from the self-guided BFO programme in which all users, regardless of their actual level of engagement with the system, were included in our analysis. Despite using modern prediction modelling methods, we were unable to accurately predict a range of engagement metrics from baseline assessment data. This suggests that it is not possible to predict who will engage more with the BFO programme from clinical information at first access.

Multiple unmeasured, and potentially unmeasurable, factors may make prediction of user engagement challenging. Difficulty in predicting user engagement is likely to be exacerbated in individuals struggling with substance use whose often unstructured lifestyle, and multiple areas of clinical complexity, likely interfere with engagement. Given the lack of prediction accuracy in this study, triaging new addiction service clients for BFO use on the basis of their baseline assessment data may exclude individuals who may engage and possibly benefit from the programme if they were introduced to it.

This research has some limitations. First, our metrics of engagement were behavioural, and do not reflect possible cognitive or emotional involvement of users with the BFO programme. Further, engagement does not equate to benefit, which may be achieved after minimal engagement. However, by including continuous engagement variables, we have attempted to reflect that beneficial engagement can have individually different outlooks and that the binarization of it on the basis of often arbitrarily defined “minimal engagement” often ignores this. In its account for the number of temporally distributed user events, our approach also agrees with recent conceptualisations of engagement as “continuing to come back” to a DI [10]. Our attempt 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 and the examination of other DIs is desirable but challenging due to limited access to commercially-sensitive data sets for independent researchers.

# 5. Conclusion

Early prediction of engagement, and intervention before dropout are desirable in digital mental health. Our case study of prediction modeling of engagement in digital CBT for substance use suggests that information beyond clinical baseline characteristics is necessary to achieve accurate predictions.

**References**

[1] Gan DZQ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: A systematic review and meta-analysis. Front Digit Health. 2021 Nov;3:764079, doi: <https://doi.org/10.3389/fdgth.2021.764079>

[2] 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: <https://doi.org/10.1136/eb-2018-102891>

[3] Baumel A, Kane JM. Examining predictors of real-world user engagement with self-guided eHealth interventions: Analysis of mobile apps and websites using a novel dataset. J Med Internet Res. 2018 Dec;20(12):e11491, doi: [10.2196/11491](https://doi.org/10.2196/11491)

[4] Marinova N, Rogers T, MacBeth A. Predictors of adolescent engagement and outcomes – a cross-sectional study using the togetherall (formerly Big White Wall) digital mental health platform. J Affect Disord. 2022 Aug;311:284-293, doi: <https://doi.org/10.1016/j.jad.2022.05.058>

[5] 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: <https://doi.org/10.1046/j.1360-0443.1995.9056072.x>

[6] 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: <https://doi.org/10.1176/appi.psy.50.6.613>

[7] 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: <https://doi.org/10.1023/B:QURE.0000018486.91360.00>

[8] 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: <https://doi.org/10.3109/10826084.2016.1161052>

[9] 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

[10] Torous J, Michalak EE, O’Brien HL. Digital Health and Engagement—Looking behind the measures and methods. 2020 Jul;3(7):e2010918, doi: 10.1001/jamanetworkopen.2020.10918

1 Corresponding Author: Franziska Günther, Department of Computer Science, Kilburn building, M13 9PL, University of Manchester, UK; e-mail: franziska.gunther@postgrad.manchester.ac.uk.