# Analysis of NGTS Treatment Impact (Tier 3 and 4 service users, 2018-2021)

Ben Hickman

Dr. Biman Chakraborty

BeGambleAware.org July 2022

# Table of Contents

# Executive Summary

## Introduction

This project examined Tier 3 (structured) and Tier 4 (residential care) treatment referrals across the National Gambling Treatment Service (NGTS) with a referral and completion date between 1st April 2018 and 31st March 2021. The NGTS is a network of organisations that provide confidential, free-to-access treatment and support for anyone experiencing gambling-related harms. The data were collected through GambleAware's Data Reporting Framework (DRF).

We examined the impact of treatment on changes in measures of gambling behaviour and wellbeing for 14,462 qualifying referrals with 94,966 attended appointments. We performed analysis at both the referral level and the appointment level. Gambling behaviour was measured using the Problem Gambling Severity Index (PGSI) and wellbeing was measured using the Clinical Outcomes in Routine Evaluation (Core-10) measure.

More service users were gamblers than affected others (85% vs 15%), with males representing a larger proportion of gamblers (84%) and females a larger proportion of affected others (85%). Service users had a mean age of 37 years at time of referral, with gamblers being on average 7 years younger than affected others. 24% of the referrals were for recurring treatment.

Of service users who attended at least two appointments, the majority completed scheduled treatment (81%). More affected others (88%) completed treatment than gamblers (80%). The mean length of treatment for gamblers was 12 weeks (mean of 6.6 sessions) and for affected others was 11 weeks (mean of 6.4 sessions); gamblers also missed more appointments (mean 1.7 sessions) than affected others (mean 1.2 sessions).

## Outcomes of Tier 3 treatment for gamblers

- 20% of service users were in the most severe category of Core-10 at first appointment. This was higher for female than male gamblers (29% vs 20%) and affected others (16% vs 6%). 82% of service users were above the clinical threshold at treatment start (a Core-10 score of >=11).
- 94% of gamblers were classed as problem gamblers by their PGSI score at the first attended appointment. This was higher for female than male gamblers (96% vs 94%).
- Core-10 and PGSI scores reduced considerably (10.6 and 14.1 points respectively) between first and last attended appointment, showing large initial change and more variability after around six appointments. The biggest difference in Core-10 and PGSI scores for gamblers was estimated at 14 and 11 attended appointments respectively, but the difference in scores between appointments was small and variable after appointment 6.
- PGSI behaviour domain items, such as chasing losses, decreased sharply in the first three appointments before levelling, while PGSI consequence domain items, such as financial problems and feelings of guilt, decreased more slowly and were more variable across the treatment journey.

- Number of attended appointments and treatment completion were strongly related to reductions in Core-10 and PGSI scores. There were larger reductions in PGSI in appointments where the previous appointment was attended, illustrating the negative impact of missed appointments.
- Female and older gamblers had smaller reductions in PGSI than male and younger gamblers. There was no evidence that the gender or age of the gambler associated with change in Core-10.

### Reliable and clinically significant change in Core-10 and PGSI

- 85% of gamblers (10,513) were above the clinical threshold for Core-10 at treatment start. 56% of these showed clinically significant improvement in Core-10 at treatment end. 26% showed no reliable change or reliable deterioraton in Core-10.
- 93% of gamblers (11,486) met the proxy clinical threshold for PGSI (score >=9) at treatment start. 77% showed clinically significant improvement at treatment end. 9% showed no reliable change or reliable deterioraton in PGSI.
- Gamblers who had made clinically significant improvement in PGSI were 11 times more likely to have also made clinically significant improvement in Core-10 than those who had not.
- Brief referrals of 3 or fewer attended appointments reduced the likelihood of clinically significant change in Core-10 by 66% and in PGSI by 72% compared to longer referrals.
- Female and older gamblers were less likely to have clinically significant change in Core-10 compared to male and younger gamblers.
- Gamblers with higher scores for PGSI ‘recognition of problem’ and ‘feelings of guilt’ at treatment start were more likely to show clinically significant improvement in Core-10.

### Impact of Tier 4 treatment on gamblers

- Tier 4 gamblers had higher mean PGSI scores than Tier 3 gamblers at treatment start (21.6 vs 19.5). Scores were higher across all PGSI items except ‘Feelings of guilt’.
- At treatment end, mean PGSI scores for Tier 4 gamblers were 1.8 compared to 3.1 for Tier 3, with larger reductions in PGSI behaviour and external consequences for Tier 4 service users.
- 100% of Tier 4 service users had clinically significant change compared to 87% of the group of Tier 3 service users included in the Tier 4 analysis. This was likely due to larger reductions in the PGSI external consequences (health problems, criticism from others and financial problems).

### Impact of treatment on affected others

- At treatment start, 81% of affected others (1,717) were above the clinical threshold for Core-10. 52% showed clinically significant improvement in Core-10 at treatment end. 29% showed no reliable change or reliable deterioraton in Core-10.
- Core-10 scores reduced considerably (8.6 points) between first and last attended appointments, showing large initial change and then smaller more variable change after around 7 appointments.

- The biggest difference in Core-10 score was attained at 15 attended appointments, but the difference in scores between appointments was small after appointment 7.
- Affected others with brief referrals were over 75% less likely to have clinically significant improvement in Core-10 compared to affected others with longer referrals.

### Brief referrals and treatment completion

- Affected others had a 67% higher likelihood of a brief referral than gamblers.
- Female service users and older service users were less likely to have brief referrals, as were service users with high baseline Core-10 depression/anxiety scores at treatment start.
- Service users who completed treatment were five times less likely to have a brief referral than those who dropped out.
- Affected others were 50% more likely to complete treatment than gamblers. Service user gender and age were not associated with treatment completion.
- High baseline Core-10 scores reduced the likelihood of treatment completion. Service users with high risk-to-self were 10% less likely to complete treatment.

### Conclusions

Our analysis found that when service users engaged in treatment and attended more than one appointment, they were more likely to have improved Core-10 and PGSI scores and make clinically significant improvement. Other factors that impacted the likelihood of improvement related to both referrals (attended appointments, treatment completion) and the service user (gender, treatment recurrence and severity at start of treatment).

There is considerable variation in treatment length in the NGTS. We found that service users with brief referrals (3 or fewer appointments) had lower levels of clinically significant change for both PGSI and Core-10.

Affected others had a lower change in Core-10 score than gamblers but were no less likely than gamblers to achieve a clinically significant change in Core-10 by treatment end. Affected others were more likely to have brief referrals and to complete treatment than gamblers.

Outcome reporting should be clear on who and what is included in any calculations and should aim to use clinically significant change as a standard. Where the PGSI is reported at treatment end, we suggest using reliable change in PGSI scores (i.e. change in score of >=5) instead of crossing screening thresholds (i.e. moving below a score of 8). We also suggest development and testing of a threshold for clinically significant change, here proposed as <9.

We make several recommendations for changes to the DRF that will allow better capture and analysis of data on treatment impact.

## Glossary

| Term | Definition |
|---|---|
| Affected other | Anyone affected by another person's gambling, for example a partner, family member or friend. |
| Appointment | Any single attended or unattended treatment session. |
| Attended Appointment Order | Attended appointment order was calculated using the attendance values of each appointment within a referral and ordering them according to date. |
| Brief referral | In this report, a brief referral is any Tier 3 treatment referral that had <= 3 attended appointments. |
| Clinically Significant Change | Clinically significant change relates to the question: 'Is the end-state score better represented by a non-clinical or general population rather than a clinical population?' and is reliable change combined with being above a clinical threshold at treatment start and below a clinical threshold at treatment end. |
| Coefficient | The coefficient value represents the mean change in the dependent variable given a one unit change in the predictor variable. |
| Domain | This is a sub-set of items within an outcome measure which are inter-related with each other, for example within the PGSI four items specifically relate to gambling behaviour. |
| DRF | The Data Reporting Framework (DRF) is the nationally co-ordinated dataset for data on service users receiving treatment from the NGTS. |
| Effect size | This is measured using Partial Eta Squared, which is the proportion of variance associated with the effect. |
| NGTS | The National Gambling Treatment Service (NGTS) is a network of organisations working together to provide confidential treatment and support for anyone experiencing gambling-related harms in England, Scotland and Wales. Up to 31st March 2021, the NGTS included GamCare and its partner network, Gordon Moody, Central and North West London Foundation Trust (National Problem Gambling Clinic) NHS Northern Gambling Service, provided by Leeds and York Partnership NHS Foundation Trust |
| OR (Odds Ratio) | The odds ratio for comparing two groups (A, B) on their impact of an outcome occurring is the odds of the event occurring for group A divided by the odds that it occurs for group B. This could also be thought of as the likelihood of an outcome for A compared to B. |
| Reliable Change | Reliable change relates to the question: 'Has the service user's score changed sufficiently to be confident that the change is not attributable to measurement error?' and is the extent to which change falls beyond that likely based on the measurement variability of the measure (the standard error of change). |
| Referral | In this report, as in the DRF, referral is used to refer to the complete treatment journey or treatment episode of a service user. |
| Tier 2 | Typically 1-3 interventions rooted in CBT and Motivational Interviewing principles. Usually delivered by GamCare and Partner Network but may also be delivered by other third sector providers such as CAB. |
| Tier 3 | A: Following a comprehensive assessment, typically delivered by GamCare and Partner Network, 6-12 structured psychosocial/treatment interventions (though not time limited).<br>B: Typically delivered by NHS providers, treatment approach determined following assessment, usually structured psychosocial/clinical interventions for those with multiple and/or severe complex needs. |
| Tier 4 | Structured treatment that involves a period of treatment in residential care. Tier 4 residential treatment is available to all those with severe gambling related harms and expressing a preference for residential care. |
| Treatment Recurrence | The treatment recurrence indicator is self-reported data on whether a service user has had treatment for gambling previously provided by the NGTS or not (treatment naivety). |

# 1 Introduction

## 1.1 Scale of problem gambling

Over the last decade there has been ‘unprecedented growth’ in commercial gambling (Blank et al. 2021). The World Health Organization (WHO 2017, p.2) uses the same phrase dating back to the mid-1980s and attributes the growth to ‘the intersection of gambling and financial technologies, impacts of internet and mobile devices [and] the spread of gambling to traditionally non-gambling settings’.

There has also been a broadening of possible gambling activities. Cooper-Somonini et al. (2021) report increased demand at the National Problem Gambling Clinic in London for support for excessive trading and stock market investments. The World Health Organization (2022) reports on the convergence on some platforms of gaming and gambling, which ‘may give rise to migration from games to gambling and co-occurrence of the two disorders’.

In Great Britain, approximately 0.5% of adults are estimated to be problem gamblers (Public Health England, 2021), although this figure varies according to data source and method. It is estimated that for each problem gambler, between six to ten further people (likely to be family members, friends and employers) may be negatively affected (Goodwin et al. 2017; Nash et al. 2018).

## 1.2 The National Gambling Treatment Service

In response to these numbers and growing evidence of personal and social harms related to gambling, the NHS Long Term Plan (NHS England, 2019) extended commitment to the funding of gambling treatment services.

The National Gambling Treatment Service (NGTS) is a network of organisations working together to provide confidential treatment and support for anyone experiencing gambling-related harms. It is free to access across England, Scotland and Wales. Service users may be either gamblers or those who are impacted by someone else’s gambling (affected others). The NGTS is commissioned by GambleAware, an independent grant-making charity that takes a public health approach to reducing gambling harms.

When an individual who is experiencing difficulties with gambling (whether a gambler or an affected other) contacts a provider in this network, the provider works alongside others through referral pathways (e.g. GPs) to deliver the most appropriate package of care. In 2021, the NGTS included the following organisations, with details of the services they provide listed below.

GamCare and its partner network:

- Online treatment supported by regular contact with a therapist, which can be accessed at a time and place convenient for the service user over the course of eight weeks.

- One-to-one face-to-face, online and telephone therapeutic support and treatment for people with gambling problems as well as family and friends who are impacted by gambling.
- Group based Gambling Recovery Courses delivered face-to-face or online for between six to eight weeks.

Gordon Moody:

- Residential Treatment Centres – two unique specialist centres, providing an intensive residential treatment programme for men with a gambling addiction over a period of 14 weeks.
- Recovery Housing – specialist relapse prevention housing for those who have completed the treatment programmes requiring additional recovery support.
- Retreat & Counselling Programme – retreat programmes for women-only-cohorts and men-only-cohorts which combine short residential stays with at-home counselling support.

Central and North West London NHS Foundation Trust (London Problem Gambling Clinic):

- Treatment for gambling problems especially for people with more severe addictions and those with co-morbid mental and physical health conditions, those with impaired social functioning, and those who may present with more risk, such as risk of suicide.

NHS Northern Gambling Service, provided by Leeds and York Partnership NHS Foundation Trust:

- Treatment for gambling problems especially for people with more severe addictions and those with co-morbid mental and physical health conditions, those with impaired social functioning, and those who may present with more risk, such as risk of suicide.

The collection of data on service users receiving treatment from the NGTS is managed through a nationally co-ordinated dataset known as the Data Reporting Framework (DRF). Providers use a wide range of interventions, with Tier 3 typically focusing on providing 6-12 structured psychosocial treatment appointments. The DRF attempts to collect the type of intervention used at each appointment but this is almost exclusively categorised as general “counselling”.

As part of the assessment and treatment process undertaken by providers within the NGTS, two standardised measures are typically collected from gamblers at initial assessment and at each attended appointment: the Core-10 and the Problem Gambling Severity Index (PGSI).

The Core-10 is validated for use as a clinical outcome measure (Barkham et al, 2013), however the PGSI is typically only used as a population screening tool (Caler et al, 2016), with validity testing of the PGSI (for example, Orford et al, 2010 and Miller et al, 2013)[1] having been undertaken within that

[1] Validity testing of the PGSI as a population screening tool has included both factor analysis (Orford et al, 2010) and comprehensive item response analysis (Miller et al, 2013).

context. Although it was not designed as an outcome measure for treatment, the PGSI is currently the only gambling specific measure being collected within the NGTS through the DRF (GambleAware, 2020).

## 1.3 Project Summary

A recent study of GamCare service users (Hickman et al, 2021) found improvement in Core-10 and PGSI scores between the start and end of treatment. One of the identified limitations of this research was that it only examined measures at the start and end of treatment and as such ignored much of the client treatment journey.

This project aimed to build on this research and examine the impact of treatment on PGSI and Core-10 measures over the entire treatment journey for the whole of the NGTS. To do this we analysed individual appointment data from the DRF for referrals completed within a three-year period between 1st April 2018 and 31st March 2021.

### Primary Research Questions

1. How do PGSI and Core-10 scores relate to treatment attendance?
2. What is the most effective minimum number of attended treatment appointments?
3. Do individual items in the PGSI and Core-10 relate to treatment at different rates?

### Secondary Research Questions

4. Does the number of days between appointments relate to treatment effectiveness?
5. Does the proportion of missed treatment appointments relate to treatment outcomes?
6. How does treatment naivety (first time in treatment) relate to treatment on Core-10 and PGSI measures?
7. What differences are observed in treatment across different service user groups (e.g. gamblers compared to affected others, gender, age)?

## 1.4 Gambling related outcome measures

Recent changes in the clinical definition of problem gambling are reflected in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (DSM). With the publication of DSM–5 in 2013, what had been termed 'pathological gambling' in the earlier DSM–IV became 'gambling disorder' and was classed as a substance-related and addictive disorder rather than an impulse-control disorder. A similar change was made in the most recent edition of the WHO's International Classification of Diseases, ICD-11 (implemented from 1 January 2022).

The DSM–5 criteria for gambling disorder require an individual to exhibit four or more criteria in a 12-month period with diagnoses of 'mild', 'moderate' and 'severe' related to specific numbers of criteria being met (four to five, six to seven, and eight to nine respectively). The DSM–5 diagnosis

criteria are one of many gambling related outcome measures, with the variation in use of measures a challenge when trying to make comparisons of effectiveness between one study and another, or one kind of treatment and another.

Ribeiro et al. (2021) conducted a systematic review of randomised controlled trials (RCTs) involving non-pharmacological treatments. The authors listed the outcome measures used in the 22 studies they included together with a brief description of each and the domains they cover. There were 22 measures, of which none were used in more than three of the studies. Some studies used additional measures such as the weekly amount of money spent, and the time and frequency of gambling. Of the two measures regularly collected in the UK – the Core-10 and the PGSI – the PGSI was used in one study, the Core-10 not at all. Five different measures were used to assess gambling severity.

In a similar vein, in their systematic review on measuring outcomes in gambling disorders, Pickering et al. (2018) included 34 treatment studies which used 63 different outcome measures. In their introduction the authors comment that 'Treatment outcomes for gambling disorders are defined poorly and measured inconsistently across studies' (p.411).

## 1.5 Impact of psychological therapy

A systematic review conducted by Chebli et al. (2016; in Kalbfleisch et al. 2020; see also GREO 2017a), included 16 studies of internet-based interventions for addictive behaviours, four of them for gambling. Most used cognitive behaviour therapy (CBT) and some used motivational interviewing. All of them had positive treatment outcomes. The studies on gambling found that problematic gambling behaviours and associated mental health problems were reduced after treatment and at follow-up.

Writing about treatment for the impact of problem gambling provided by the National Problem Gambling Clinic and NHS clinics, Cooper-Samonini et al. (2021) describe CBT as having been found 'helpful' in reducing gambling frequency, achieving better rates of abstinence and maintaining treatment effects at follow-up. This assessment refers to the findings of Gooding and Tarrier (2009), Ribeiro et al. (2021) and Toneatto and Millar (2004).

Chebli et al. (2016) undertook a systemic review in which all 16 studies used internet-based CBT with clinical assistance via chat or email and reported improvements in gambling behaviours lasting up to three years after the intervention. Van der Maas et al. (2019) conducted a scoping review on internet-based interventions for problem gambling. They found that CBT was the most common internet-based intervention and that such interventions more generally were effective in reducing problem gambling scores and gambling behaviours.

Cooper-Samonini et al. (2021) refer to an alternative approach to CBT researched at the National Problem Gambling Clinic, brief relational psychodynamic therapy (see also GREO 2019a; Mooney

et al. 2019). The authors of the study found this to be an effective approach for individuals with co-occurring mental health problems or other addictions where CBT or a 12-step approach had not been effective. As well as treating problem gambling with some success, it also reduced depression and anxiety.

The literature refers to a range of interventions described as “Motivational”. Kalbfleisch et al. 2020 say that the evidence supporting motivational interventions is limited and the results tend not to be robust or maintained. Blank et al. (2021) include in their mapping review a meta-analysis by Yakovenko et al. (2015) with five studies published between 2001 and 2009 using motivational interviewing, mostly one to one. They found a significant reduction in the frequency of gambling but not in expenditure on gambling at six-month and 9–12-month follow-ups. In their rapid evidence review, Kalbfleisch et al. (2020) include a study by Boudreault et al. (2018; see also GREO 2018) of an 11-week self-help treatment based on use of a workbook for ‘at-risk and pathological gamblers’. A component was inclusion of three motivational interviews conducted by phone. The authors reported that the intervention led to a significant reduction in gambling behaviours and consequences, and they suggested that it was the motivational interviews which contributed to the success of the treatment.

The effectiveness of motivational interviewing in combination with other therapies was also the subject of a systematic review conducted by Petry et al. (2017) and is included in the reviews of both Blank et al. (2021) and Kalbfleisch et al. (2020). Based on 21 trials, the authors concluded that there are benefits – in terms of less time and money spent on gambling – from CBT alone and from CBT combined with motivational interviewing but not from motivational interviewing alone.

Marchica and Derevensky (2016) undertook a systematic review including six studies using personalised feedback intervention (PFI), an approach they say has proven successful in reducing alcohol consumption. The studies all reported a reduction in aspects of gambling behaviour, some of which were significant. The authors of the systematic review describe PFI as a promising intervention despite the limited evidence.

## 1.6 The importance of treatment length

The literature refers to some treatments as ‘brief’ interventions. Blank et al. (2020), in their mapping review, include two definitions of ‘brief interventions’: single-appointment interventions and those with a maximum of three appointments. Assessments of effectiveness lean towards the view that outcomes from brief interventions can be as effective as from longer ones.

Peter et al. (2019) conducted a meta-analysis of brief personalised feedback interventions for problematic gambling, which found that in single-appointment interventions the strongest predictor of positive effect in the short term was having an educational element.

In a systematic review of 22 non-pharmacological treatments, Ribeiro et al. (2021) concluded that, in relation to CBT, longer interventions seem to be no more beneficial than shorter ones. In all the studies where the duration was specified there were significant improvements to outcomes. Parker and Bauermann's (2015) meta-analysis concluded that the effectiveness of psychological interventions was unrelated to either the number of appointments or the number of hours.

## 1.7 The importance of treatment completion

While there are frequent references in the literature to the high dropout rate in treatment for problem gambling, there is little evidence on differences in outcomes between those who completed treatment and those who dropped out. In their systematic review of RCTs, Ribeiro et al. 2021 (referring to Potenza et al. 2019) commented that two out of three people with gambling disorder who seek treatment give up before the end of the scheduled treatment programme. Hickman et al. (2021), in their analysis of GamCare treatment service users, reported that almost a third do not complete the planned treatment programme. Similarly, Kalbfleisch et al. (2020), commenting specifically on CBT, say that treatment drop out is prevalent and that this affected the findings of many of the studies in their rapid evidence review.

The National Gambling Treatment Service (GambleAware, 2021) reported in its annual statistics a lowering of the dropout rate in comparison with five years earlier, from 35% to 20%. A greater proportion of service users who completed treatment were reported to show improvements in PGSI scores (92%) than those who dropped out (60%), and in Core-10 scores (88% of those who completed treatment improved compared to 52% of those who dropped out).

## 1.8 Concluding observations

Much has been written about problem gambling and its increasing prevalence, however, there is a marked sense that the literature does not provide sufficient specificity about what treatments and what aspects of treatments are effective in addressing the problem. Meaningful comparisons are further complicated by the profusion of outcome measures in use.

The literature covers a range of types of treatment (psychological, behavioural, motivational) and combinations of treatment, and a range of ways in which they can be delivered (in person, online, via phone) but it also points to unevenness in what is researched. The treatments provided by the NGTS suffer the same variance in combinations of treatment and delivery modes.

It would be beneficial to examine the impact of shorter (3 or less appointments) treatment referrals. It should be noted that there are also Tier 2 'brief interventions' undertaken by NGTS providers which are not included within the DRF scope at present.

This project provides an opportunity to examine in greater depth the impact of treatment completion on service user's reported outcomes. GambleAware (2021) reported in its annual NGTS statistics

greater improvement rates in those who complete treatment. However, no information was provided on how far through treatment individuals were when they dropped out. These figures may therefore be skewed by service users who only have a single session of treatment and then never return. This project examines whether treatment completion has an impact on outcome scores for service users who engage with treatment (i.e. attend more than one session).

## 1.9 Report structure

The remainder of the report has the following structure:

- Section 2 provides the full methodology for the project including scope, data sources and cleaning, definitions of measures and analysis methods.
- Section 3 provides an overview of the referrals included in the analysis.
- Section 4 summarises the findings related to the impact of treatment on gamblers' PGSI and Core-10.
- Section 5 summarises the findings related to the impact of treatment on affected others' Core-10 and compares outcomes between gamblers and affected others.
- Section 6 provides discussion of the findings, conclusions, limitations and recommendations.

# 2 Methodology

## 2.1 Data Reporting Framework

Data on service users receiving treatment from the NGTS is managed through a nationally co-ordinated dataset known as the Data Reporting Framework (DRF). Individual treatment providers collect data from service users and their treatment through bespoke case management systems. The definitions provided by the DRF are incorporated into each of these systems, with regular submissions of pseudonymised data made to a centrally managed system. This system combines the data from each service provider into a single dataset.

Data are collected within four separate tables:

- **Person** which contains service user characteristics including demographics, location, and relationship and employment status.
- **Gambling history** which contains details of a service user's gambling behaviour and related history, including impacts of their gambling behaviour such as time and money spent on gambling.
- **Referral** which contains information about a service user's treatment referral (also known as an episode), such as referral source, provider, start and end dates and whether they have previously received treatment.
- **Appointment** which contains details including the PGSI and Core-10 scores from each appointment, where an appointment is defined as a scheduled interaction with a service user with the objective of improving the overall health of the service user.

The DRF constitutes a co-ordinated core data set, collected to provide consistent and comparable reporting at a national level. Some minor differences exist in data collection between agencies, such as the addition of supplementary categories in individual fields or in the format of collected data. These are reformatted or recoded at a national level to ensure consistency.

It is important to note that throughout this report "referral" will be used to refer to the complete treatment journey or episode of a service user. "Appointment" will be used to refer to a single treatment session.

## 2.2 Scope

This project examined all Tier 3 (structured) and Tier 4 (residential care) treatment referrals across the NGTS that had at least two attended appointments and qualified for inclusion in the DRF with a referral and completion date between 1st April 2018 and 31st March 2021. These dates were chosen to ensure that the data would be available within the DRF in the appropriate project timescales and avoided definitional changes that applied from April 2021 onwards. All appointments between the referral and completion dates of the referral were included.

## 2.3 Data source

The project used data from three tables of the DRF. The exact data used is listed in Table 1 below.

***Table 1: DRF data items included in analysis***

| | | |
|---|---|---|
| **Referral Table** | | |
| - Care Plan Number | - Date referral received | - End reason |
| - Local Service User Identifier | - Referral reason | - End date |
| - Provider Code | - Previous treatment for Gambling | |
| **Person Table** | | |
| - Care Plan Number | - Provider Code | - Local Service User Identifier |
| - Gender Identity | - Age at Referral | |
| **Appointments Table** | | |
| - Care Plan Number | - Appointment date | - PGSI score |
| - Local Service User Identifier | - Attendance | - CORE-10 score |
| - Provider Code | - Appointment purpose | |

Alongside the data obtained directly from the DRF, we also requested additional breakdowns of all PGSI and Core-10 measures into the composite item scores directly from each of the four providers. The PGSI is made up of 9 items and the Core-10 is made up of 10 items. Full details of the items and domain groupings within both measures are available in section 2.6. Three of the providers were able to provide full breakdowns of both the Core-10 and the PGSI. These data were then linked to the DRF using the Care Plan Number, Local Service User Identifier, Appointment Date, and the appropriate total PGSI or Core-10 Score.

Following initial inspection of the DRF and discussions with Gordon Moody, it became clear that the DRF data from Gordon Moody (which would mainly relate to Tier 4 treatment) was not compatible with the DRF data from Tier 3 treatment providers. This is because the PGSI and Core-10 measures in the DRF are linked to each appointment, but in the 14-week residential care service provided by Gordon Moody there are no defined “appointments”. Service users instead receive intensive treatment and care on each day of their 14-week residential stay and are isolated from the outside world in a way that makes continuous collection of measures, particularly the PGSI, inappropriate. This also meant that the “appointments” recorded in the DRF by Gordon Moody for residential care normally equated to a single appointment for each week the service user stayed in residential care. Service users in Tier 3 treatment could potentially have more than one appointment (including outcome scores) per week. It was decided to analyse data from Gordon Moody separately to the main Tier 3 analysis, details of which can be found in section 2.9.

## 2.4 Data Cleaning

There were 26,459 referrals included in the initial DRF data with a start date between 01/04/2018 and 01/04/2021. These referrals had a total of 170,368 appointments attached, as shown in Table 2.

*Table 2: DRF referral data with a start date between 01/04/2018 and 01/04/2021*

| Agency | Referrals | % of total referrals | Appointments |
|---|---|---|---|
| GamCare | 23,629 | 89.3% | 155,092 |
| Gordon Moody | 1,609 | 6.1% | 4,945 |
| London Problem Gambling Clinic | 612 | 2.3% | 4,610 |
| NHS Northern Gambling Service | 609 | 2.3% | 5,721 |
| **Total** | **26,459** | | **170,368** |

The following cleaning rules were agreed with GambleAware and applied to the data once the 1,609 referrals to Gordon Moody had been excluded:

### Appointments data

- Remove appointments with a purpose of "follow-up after treatment" as they fall outside the scope of this project.
- Combine duplicate unattended appointments that occurred on the same day into a single unattended event. This is to reduce the impact of differences in how missed appointments are recorded by the various providers.
- Combine duplicate attended appointments that occurred on the same day only when measures are the same from both appointments, keep all other appointments that occurred on the same day. 98% of occurrences of two appointments in one day are when an assessment and treatment has occurred on the same day but been entered as two separate appointments. By only combining when they have the same measure scores, we ensure that no data is lost while reducing differences in how data are recorded locally.

### Referrals data

- Exclude any referrals that had no end date or an end date after 01/04/2021, as these would be outside the scope of the project (1,932 referrals excluded).
- Exclude any service users with a referral reason of "At Risk of a Gambling Problem" (272 referrals excluded). The sample for "at risk" service users is not big enough to treat it as a separate group so it is excluded from this analysis as it would seem inappropriate to combine 'at risk' service users into the 'gamblers' client type.
- Exclude any referrals which had less than two attended appointments, with appropriate measures attached, excluding follow-up appointments (6,927 referrals excluded). Only one PGSI and Core-10 is recorded against each appointment within the DRF, which makes assessment of change impossible for any referrals which have only a single attended appointment, and the appointment must have more than one available measure for impact to be assessed.
- Exclude any referral with a treatment end reason of any of the following: Deceased (Assessed & treated), Not suitable for service – no action or referral back, Not suitable for

service – signposted elsewhere, Referred on (Assessed & treated), Referred on (Assessed only), or Treatment Declined (total of 231 referrals excluded). This project is assessing the impact over the whole treatment journey so should include only service users who have finished their treatment (either through completion or discontinuation). If, for example, a service user dies during treatment or is referred on to another treatment service, their treatment journey has not ended and so their data should not be included in this analysis.

- Exclude service users who have more than one referral recorded as occurring at the same time (14 referrals excluded). There are a small number of service users who have more than one referral recorded as occurring at the same time. This is probably due to a data entry issue but they have been removed from the analysis to ensure integrity of the data.
- Exclude referrals which were missing measures (including the individual item breakdowns) from the first attended appointment (715 referrals). There is no way to reliably infer what had occurred at the first session and what impact this may have had on their initial measure scores. As such these will be excluded from the analysis.
- There were 462 referrals where the final attended appointment had no available measure. Instead of removing them from the analysis we assessed total attended measures up to the point that the last available measure is available, as there is no way to assess the impact of any attended appointments after this point. We will include a variable in the referral level regressions that includes the total number of attended measures after the final measure was taken to assess its impact.
- Exclude referrals which were missing the individual item breakdown from measures at their last attended appointment (56 referrals). To ensure that we can compare the impact of treatment across measure domains we will need individual item scores at both the start and end of treatment.

The total number of referrals excluded on the basis of the above rules was 11,756.[2]

[2] It is worth noting that the number of referrals affected, as listed under each process above, is dependent on its order within the overall process. So, for example, the number of referrals affected by the removal of “at risk” gamblers is higher than it would be if it came later in the process and is lower than if it came before the date exclusions. This does not affect the final number of referrals.

### Attended appointment outliers

Initial analysis highlighted that while most service users (97.2%) had 15 or fewer attended appointments, a small minority of service users (n=137) continued in treatment for more than 20 attended appointments (with the maximum being 56 attended appointments).

Of the 137 referrals which lasted longer than 20 attended appointments, 26 (19.0%) of them were for affected others. While this was a higher proportion than for the remaining data (14.6%) it was not significantly[3] different. Service users with >20 attended appointments had higher[4] Core-10 scores at the first attended appointment (Mean=20.2, Standard Deviation = 7.7) compared to referrals with =<20 attended appointments (M=18.1, SD=8.1). PGSI scores were similar at first appointment between gamblers with >20 attended appointments (M=19.4, SD=6.2) and gamblers with =<20 attended appointments (M=19.0, SD=6.2).[5]

Including service users with >20 attended appointments resulted in the sample size for later appointments (i.e. those appointments occurring after 20+ appointments had already been attended) being too small to make meaningful conclusions (e.g. data on the 30$^{th}$ attended appointment was based on only 33 service users). The decision was taken to exclude service users who attended more than 20 treatment appointments from the analysis. For service users to stay in treatment for this long they are likely to have requirements for support outside of what would normally be expected, which is supported by their higher Core-10 scores at treatment start. It would be outside the scope of this project to examine these in more detail, as this would require additional information on the service users beyond what we have available, but this will be recommended in the further work recommendations in section 6.7.

This resulted in the further exclusion of 137 referrals. The mean attended appointments for the excluded group was 26.9, compared to an overall mean of 6.5, with the mean treatment length for the excluded group being 8.9 months.

After these exclusions there remained 12,434 referrals for gamblers and 2,132 for affected others. Of these, 3,528 referrals had only 2 or 3 attended appointments; this equates to nearly one quarter (24%) of the qualifying referrals. In Quilty et al. (2019) these were considered to be brief interventions. For certain elements of the analysis, we will compare any differences in these shorter referrals to the remaining, longer, treatment referrals.

[3] Chi squared test X2(1, 14,599) = 2.063, p = .151.

[4] T test t(14,597) = -2.924, p = .003

[5] T test t(12,457) = -.696, p = .486

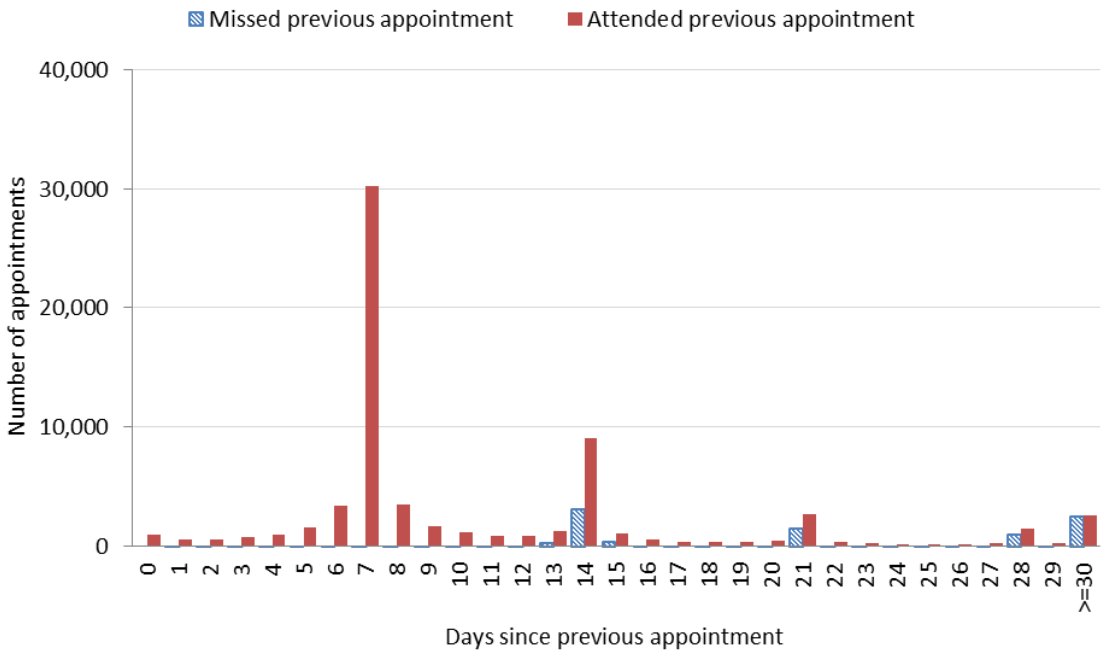
### Days between attended appointments

The majority (77%) of appointments occurred within two weeks of the previous appointment. However, within the DRF data there were also appointments that occurred months after the previous attended appointment, with 6% of appointments having a gap of 30 days or more since the previous attended appointment.

These appointments may have been the start of a new referral following an extended period of non-contact with a service user, but there was no way to identify this from legitimate extended gaps in treatment. Any analysis of days between appointments would be skewed by these (possibly anomalous) appointments. As these appointments could be in the middle of a referral, it was not possible to exclude them without removing more (possibly legitimate) data than was proportionate.

The other key reason for longer gaps between attended appointments was non-attendance of the previous appointment. For appointments that occurred within 13 days or less of the previous appointment, only 2% of them had missed a planned appointment. For those appointments with 14 days or more since the last attended appointment, 34% had missed a planned appointment. This increased to 50% of appointments with a gap of 30 days or more since the last attended appointment (Figure 1).

***Figure 1: Count of attended appointments by days since previous attended and planned appointment***



The 11,723 attended appointments which were preceded by a missed appointment had a mean days-since-last-attended-appointment of 25.0, compared to 11.4 for appointments preceded by an attended appointment.[6]

As missed appointments had such an impact on the days between appointments, we decided to include variables on planned appointments that were missed into the referral level analysis and a variable on whether the previous appointment was missed to the appointment level analysis. This allowed us to account for this impact within our modelling framework.

## 2.5 Missing data

Missing data profiles needed to be provided separately for gamblers and affected others as PGSI is not routinely collected for affected others.

Due to the data cleaning rules that were required to ensure that the data was suitable for analysing the impact of treatment (particularly the requirement that referrals should have at least two measures, with a measure available at the first attended appointment), the only missing values at referral level for gamblers are in service user age and gender (n=104). As outlined in the missing data profiles provided in Annex 1, there was no evidence of patterns in the missing values.

We used complete case analysis, also known as listwise deletion (LD), which utilizes only the cases in a data set for which there are no missing values on any of the variables. To ensure consistency in reporting we removed the 104 referrals which were missing data on service user age or gender before analysis to create complete and consistent data at the referral level.

### Missing Appointment level data

Assessing whether there are patterns in the missing data at appointment level is complicated by the fact that we would expect measure scores to be missing for appointments that were missed; appointments that were missed would automatically have missing attended values. Therefore, patterns in the missingness are expected in these data.

We therefore looked only at missing data for attended appointments; our justification for this is that these are the main components of the linear mixed model. An indicator of whether the previous appointment was missed is included in the missing data profile as this is how unattended appointments are accounted for in the mixed model. The previous attendance indicator was missing for many first appointments as the majority (11,717) of first appointments were not preceded by any appointment, missed or attended. To account for the uniqueness of first appointments, we used the change in PGSI or Core-10 (and domains) in each appointment compared to the first appointment scores as dependent variables and excluded the first attended appointments. The first appointment

[6] T test $t(13178) = 75.2$, $p = <.001$

is not included as an event in the model as its scores are used as a baseline for the entire treatment journey. This allows us to identify when the change in PGSI or Core-10 is the greatest in comparison to the baseline scores.

We ran separate missing data profiles for PGSI (for gamblers only) and Core-10 (for gamblers and affected others). Missing data patterns were observed for the PGSI scores, so we imputed missing values for PGSI (and for consistency Core-10 for gamblers and affected others). The multiple imputation model used in this case was linear regression (Rubin, 1987); full details of the imputation method are provided in Annex 2.

### Summary

Following the application of all the above cleaning rules, the decision to separately analyse Gordon Moody data, the exclusion of referrals with more than 20 attended appointments and the exclusion of referrals with missing gender or age data, there remained 14,462 qualifying referrals, with 118,865 appointments and 94,966 attended appointments with measures (Table 3). As first attended appointments are incorporated as a baseline and therefore not included as an appointment, the final number of attended appointments included in the model will be equal to the total attended appointments minus the number of first appointments (which is equal to the number of referrals as each service user only has one first attended appointment).

*Table 3: Qualifying referrals from providers following data cleaning*

| | Referrals | Appointments | Attended Appointments with Measures | Attended Appointments in Mixed Model |
|---|---|---|---|---|
| Affected other | 2,114 | 15,950 | 13,435 | 11,321 |
| Problem gambler | 12,348 | 102,915 | 81,531 | 69,183 |
| Total | 14,462 | 118,865 | 94,966 | 80,504 |

## 2.6 Core-10 items and domains

CORE-10 (Clinical Outcomes in Routine Evaluation) measures wellbeing indicators. It consists of ten items. Items are assessed on a five-point scale: not at all (0), only occasionally (1), sometimes (2), often (3), and most or all of the time (4); with total scores ranging between 0 and 40. All responses relate to the previous week.

| # | Summary | Item | Domains: Clusters | Analysis Cluster |
|---|---|---|---|---|
| 1 | Anxiety | I have felt tense, anxious or nervous | P: Anxiety | Depression/ Anxiety |
| 2 | Support | I have felt I have someone to turn to for support when needed | F: Close Relationship | Functioning |
| 3 | Able to cope | I have felt able to cope when things go wrong | F: General Functioning | Functioning |
| 4 | Talking to others | Talking to people has felt too much for me | F: Social Relationships | Functioning |
| 5 | Panic | I have felt panic or terror | P: Anxiety | Depression/ Anxiety |
| 6 | Plans to end life | I made plans to end my life | R: Risk | Risk to Self |
| 7 | Difficulty sleeping | I have had difficulty getting to sleep or staying asleep | P: Physical | Depression/ Anxiety |
| 8 | Despair | I have felt despairing or hopeless | P: Depression | Depression/ Anxiety |
| 9 | Unhappiness | I have felt unhappy | P: Depression | Depression/ Anxiety |
| 10 | Unwanted images | Unwanted images or memories have been distressing me | P: Trauma | Depression/ Anxiety |

Domains: P = Problems; F = Functioning; R = Risk

For the purposes of our analysis we have grouped the items into the following analysis domains:

1. Functioning
2. Anxiety/Depression/Trauma (which equates to problems domain in Core-10)
3. Risk to self

As the risk to self-domain is based on a single item, we have not used this in any of the analysis. These are based on the main domains used to develop the Core-10 from the Core-OM (Connell & Barkham, 2007).

**Scoring instructions**

| Core-10 Category | | Score |
|---|---|---|
| **Non-clinical** | **Healthy** | 0-5 |
| | **Low-Level** | 6-10 |
| **Clinical** | **Mild** | 11-15 |
| | **Moderate** | 16-20 |
| | **Moderate to Severe** | 21-25 |
| **Severe** | **Severe** | 26-40 |

## 2.7 PGSI items and domains

The PGSI measures problem gambling indicators. It consists of nine items. Items are assessed on a four-point scale: never (0), sometimes (1), most of the time (2), almost always (3), with total score ranging from 0 to 27. The screening relates to the previous 12 months but in the NGTS typically relates to time since the last appointment.

| # | Summary | Item | Domains | Analysis Cluster |
|---|---|---|---|---|
| 1 | Bet more than can afford | Have you bet more than you could really afford to lose? | Behaviour | Behaviour |
| 2 | Tolerance | Have you needed to gamble with larger amounts of money to get the same feeling of excitement? | Behaviour | Behaviour |
| 3 | Chasing losses | When you gambled, did you go back another day to try to win back the money you lost? | Behaviour | Behaviour |
| 4 | Borrowing money | Have you borrowed money or sold anything to get money to gamble? | Behaviour | Behaviour |
| 5 | Recognizes problem | Have you felt that you might have a problem with gambling? | Consequence | Recognition and Guilt |
| 6 | Health problems | Has gambling caused you any health problems, including stress or anxiety? | Consequence | External Consequence |
| 7 | Criticised by others | Have people criticized your betting or told you that you had a gambling problem, regardless of whether or not you thought it was true? | Consequence | External Consequence |
| 8 | Financial problems | Has your gambling caused any financial problems for you or your household? | Consequence | External Consequence |
| 9 | Unhappiness | Have you felt guilty about the way you gamble or what happens when you gamble? | Consequence | Recognition and Guilt |

For the purposes of our analysis, we have grouped the items into the following analysis domains:

1. Behaviour
2. Recognition and Guilt
3. External Consequences

These are based around the two domains of the PGSI (Ferris and Wynne, 2001) but distinguish between those consequences that are based on a person's feeling ('recognition of problem' and 'feelings of guilt') and those based on external factors ('health problems', 'criticism by others', 'financial problems'). Previous research (Hickman et al, 2021) suggested these items respond

differently during treatment. We conducted analysis combining Recognition and Guilt and External Consequences into the main Consequence domain (details are provided in section 2.10).

**Scoring instructions**

| PGSI score categories | Score | Description |
|---|---|---|
| **Non-problem gambler** | 0 | Gamblers who gamble with no negative consequences |
| **Low-risk gambler** | 1-2 | Gamblers who experience a low level of problems with few or no identified negative consequences |
| **Moderate-risk gambler** | 3-7 | Gamblers who experience a moderate level of problems leading to some negative consequences |
| **Problem gambler** | =>8 | Gambling with negative consequences and a possible loss of control |

## 2.8 Defining reliable and clinically significant change

In the Core-10 user manual (Connell & Barkham, 2007), the authors define two concepts related to the impact of treatment: reliable change and clinically significant change. Reliable change relates to the question: 'Has the service user's score changed sufficiently to be confident that the change is not attributable to measurement error?'. Clinically significant change relates to the question: 'Is the end-state score better represented by a non-clinical or general population rather than a clinical population?'

Reliable change is the extent to which change falls beyond that likely based on the measurement variability of the measure (the standard error of change), and is based on the following formula:

$$SD_1\sqrt{2}\sqrt{(1-r)} * 1.28$$

Where $SD_1$ is the standard deviation of the baseline observations and r is the reliability (coefficient alpha) of the measure.[7]

According to the testing values set out in the Core-10 manual (ibid.) ($SD_1$ = 7.9 r =.82), the Core-10 has a reliable change figure of 5.9 which is rounded up to 6 for ease of measurement (note this is a conservative approach). The recommendation in the Core-10 manual is therefore that the service user must improve by 6 or more points on Core-10 from pre- to post-therapy for us to be confident that they have made reliable improvement.

Core-10 scores at first attended appointment in the NGTS dataset were tested in the same way and provided very similar figures (N = 14,462, $SD_1$ = 8.1, r =.81) giving a reliability measure of 6.3, with gamblers (N = 12,348, $SD_1$ = 8.1 r =.81) = 6.3 having a slightly higher figure than affected others (N = 2,114, $SD_1$ = 7.7 r =.81) = 6.1. As the reliable change figure relates to whether the change is

[7] Multiplying by 1.28 (as opposed to the traditional value of 1.96) yields a value which is unlikely to occur more than 10% of the time due to the unreliability of the measure alone and is recommended for use with short outcome measures (Wise, 2004).

attributable to measurement variability, we wanted to ensure it related as closely as possible to the population in question to ensure we minimised measurement error. We decided to use the reliable change figure for the NGTS data as opposed to the one from the Core-10 manual, which is based on a smaller sample of both general population and people going through primary care counselling. The reliable change figure we used for the Core-10 was 7, which is a more conservative approach and ensured the analysis more accurately reflects the clinical population within the NGTS data.

Clinically significant change is defined in the Core-10 manual as reliable change (so change >=7 in the NGTS) from a pre-therapy score of 11 or above to a post-therapy score of 10 or below. For service users whose initial Core-10 score is already 10 or below it is therefore not possible for them to achieve clinically significant improvement. For example, someone who goes from a score of 20 to 13 shows reliable improvement, but not clinical improvement, whereas someone who goes from a score of 14 to 5 shows both.

As the PGSI was developed as a screening tool and not a clinical outcome measure there is no comparable data on what would count as both a reliable and clinically significant change. We can calculate the reliable change from the PGSI data at first attended appointment from the NGTS following the formula used in the Core-10, (N = 12,348 $SD_1$ = 6.2 r =.85) giving a reliability measure of 4.4 (rounded up to 5 for the purposes of scoring individual service users).

While the PGSI does not have a clinical/non-clinical threshold, it does have a threshold for ‘problem gambling’; a score of 8 or more. One of the issues with using this as a clinical threshold is that existing research (Hickman et al, 2021) has highlighted that gamblers in treatment often maintain higher scores for guilt and problem recognition than the general population. One way around this may be to examine only the behaviour domain. In Holtgraves (2008) the mean scores for PGSI items are provided for respondents broken down by their PGSI category. The mean score for low-risk respondents for behaviour domain PGSI items was 0.7 while for moderate risk gamblers it was 2.2. In both cases, behaviour domain score made up around 50% of the total PGSI score. We therefore propose two possible proxies for a clinical threshold:

1. Anyone with a PGSI score of 8 or more at treatment start and less than 8 at treatment end (accepting this may be impacted by scoring on guilt and problem recognition).
2. Anyone with a score of 4 or more on behaviour items only at treatment start and less than 4 on behaviour items at treatment end.

Both of these proposed clinical thresholds would also require reliable change in the total PGSI score. So for the behaviour based threshold, if a service user had a pre-treatment PGSI score of 9 and a behaviour score of 6 and an end total score of 5 and a behaviour score of 3, this would not count as clinically significant change as their overall score has not changed by >=5. Being over the thresholds at treatment start is particularly important in relation to the behaviour domains as service

users may have abstained from gambling before they entered treatment which would significantly impact their behaviour scores.

Both proxies were used and a recommendation made about possible usage going forward (which will consider the potential practical complications of using individual PGSI domains as opposed to total score).

The thresholds for reliable and clinically significant change in both Core-10 and PGSI are summarised in Table 4 below.

***Table 4: Qualifying rules for reliable and clinically significant change in PGSI and Core-10***

| Measure | Reliable Change | Clinically Significant Result (requires reliable change) | |
|---|---|---|---|
| | | At Start | At End |
| Core-10 | Change >=6 | Score >=11 | Score <11 |
| PGSI | Change >=5 | Score >=8 | Score <8 |
| PGSI Behaviour | Change in PGSI >=5 | Behaviour Score >=4 | Behaviour Score <4 |

## 2.9 Anxiety over treatment ending

The end of psychological treatment, also known as the termination phase of psychotherapy, can be defined as the part of treatment when "the therapist and the service user consciously or unconsciously work toward bringing the treatment to an end".[8] The timing of this ending phase is linked to the service user's overall satisfaction with treatment (Roe, 2007) and evidence suggests both therapist and service user can find the process distressing and a cause of anxiety (Bhatia and Gelso, 2017).

The availability of outcome measure data at each appointment, as opposed to just start and end of treatment, meant that we could assess whether anxiety over treatment ending affected outcome scores at treatment end in the NGTS. This is particularly important as the typical way to assess the impact of treatment is to examine the first and last measure scores. If the last scores are inflated by additional anxiety or stress over the ending of treatment, then they may be understating the direct impact that treatment had.

To test whether there was evidence of this we undertook paired t-tests on the Core-10 scores and the Core-10 Depression/Anxiety domain at the last and second to last appointment. We also compared the lowest Core-10 score recorded throughout the whole referral and compared it to the Core-10 score at the last appointment. We did this only for service users who had more than 3

[8] Gelso and Woodhouse, 2002, p.346

attended appointments and had completed treatment, as for anxiety over treatment ending to be a factor, the service user must know the recorded last appointment would be their last.

The paired difference equates to the score at the last appointment minus the score at the second to last appointment. As such, any positive difference may be an indication that anxiety over treatment ending is impacting the last reported measures by inflating the final score. We observed consistent negative differences for the Core-10 total and Depression/Anxiety domain, whereby scores were lower in the last than second last appointment (Table 5).

***Table 5: Pairwise t-test results comparing Core-10 score and Depression/Anxiety domain at last and second to last attended appointment***

| Client type | Core-10 domain | Paired Differences | | | |
|---|---|---|---|---|---|
| | | Mean | 95% CI | | P-value |
| | | | Lower | Upper | |
| Gambler (N = 12,348) | Core-10 Score | -2.34 | -2.44 | -2.24 | < .001 |
| | Core-10 (Depression / Anxiety) | -1.59 | -1.66 | -1.52 | < .001 |
| Affected Other (N =2,114) | Core-10 Score | -2.67 | -2.90 | -2.43 | < .001 |
| | Core-10 (Depression / Anxiety) | -1.89 | -2.06 | -1.71 | < .001 |

When we compared the Core-10 scores at last appointment with the lowest Core-10 score recorded throughout the referral we observed consistent positive differences between mean scores (Table 6). This suggests that service users generally record their lowest Core-10 scores at some point before their last appointment.

***Table 6: Pairwise t-test results comparing Core-10 score and Depression/Anxiety domain at last appointment and lowest score throughout treatment referral***

| Client type | Core-10 domain | Paired Differences | | | |
|---|---|---|---|---|---|
| | | Mean | 95% CI | | P-value |
| | | | Lower | Upper | |
| Gambler (N = 12,348) | Core-10 Score | 1.56 | 1.50 | 1.62 | < .001 |
| | Core-10 (Depression / Anxiety) | 1.03 | 0.99 | 1.08 | < .001 |
| Affected Other (N =2,114) | Core-10 Score | 1.56 | 1.42 | 1.71 | < .001 |
| | Core-10 (Depression / Anxiety) | 1.09 | 0.97 | 1.20 | < .001 |

While there is no evidence of the last recorded Core-10 being inflated due to anxiety over treatment ending, there is evidence that treatment has a non-linear impact on Core-10 scores. This means that instead of further treatment sessions consistently resulting in improvement, scores may at some point stabilise with no further improvement from additional treatment sessions. They may also fluctuate with periods of both improvement and decline. We can use the detailed appointment data available within the DRF to test whether this is the case.

## 2.10 Analysis methods

The data used in this analysis could either be at the level of the referral or appointment. Referral data includes variables linked to either specific appointments (e.g. first or last) or aggregates based on all appointments (as in total attended appointments or proportion of planned appointments missed). Appointment level data typically relates individual measures to the appointment's order within the referral (i.e. attended appointment order). This section outlines the main analysis undertaken for both types of data.

### Referral level analysis

An overview of the referrals included in the analysis is provided in Section 3, which provides descriptive statistics of key characteristics and formal statistical testing for differences between gamblers and affected others.

We used pairwise t-tests to compare the mean difference at the first attended appointment and at the last attended appointment in the Core-10 and PGSI total scores and domains scores. We used linear regression with the change in PGSI and Core-10 scores and domain scores between the first and last appointment as the dependent variable (Table 7) and the number of attended appointments when last measure was taken as the independent variable. This analysis was conducted separately for gamblers and affected others as we wanted to be able to compare how changes in PGSI might impact Core-10 for gamblers.

***Table 7: Dependent variables for linear regression analysis (change between first and last appointment)***

| Dependent variables |
|---|
| Change in PGSI score |
| Change in PGSI behaviour domain (1,2,3,4) score |
| Change in PGSI consequence (5,6,7,8,9) score |
| Change in Core-10 score |
| Change in Core-10 functioning domain (2,3,4) score |
| Change in Core-10 depression/anxiety domain (1,5,7,8,9,10) score |

Service user and referral details were used as control variables (Table 8).

*Table 8: Control variables for linear regression analysis*

| Category | Variable | Type |
|---|---|---|
| Referral | Treatment completion (1 = completed. 0 = drop out or discontinued) | Categorical |
| | Total planned appointments missed | Scale |
| | Whether referral was brief (<=3 attended appts) | Categorical |
| Service user | Known recurrence (whether service user had previously had treatment) | Categorical |
| | Referral number | Scale |
| | Gender | Categorical |
| | Age (age at referral) | Scale |

Brief referrals were defined as any referrals that had 3 or fewer attended appointments. Because this variable may be correlated with the number of attended appointments when last measure was taken, we computed Variance Inflation Factor (VIF) statistics to check for collinearity. VIF statistics are reported with the main output tables and there were no issues with multi-collinearity identified for any of the analyses.

Treatment completion is taken from the end reasons entered by the practitioner at the conclusion of treatment. As outlined in section 2.4, some referrals have been excluded from the analysis based on their treatment endings. It is also worth noting that as we are only examining referrals with two or more attended appointments, the drop-out rate will be lower than reported in the NGTS annual statistics (GambleAware, 2021) as it will not include referrals where the service user only attended a single appointment (typically an assessment) before dropping out.

The recurrence indicator is self-reported data on whether a service user has had treatment for gambling previously provided by the NGTS or not (treatment naivety). The collection of these data is imperfect as it relies on the service user informing their current provider about previous treatment, which they may be resistant to do. We recoded the data to a dichotomous variable that indicates known recurrence.

As there are service users with multiple referrals within the data, the referral number is included as a control variable. This is set to 1 for all service users with only one referral and increases by one (and is ordered by start date) for each additional referral. To model the dependency of the referrals within a service user, we fitted a linear mixed model with referral number having a random effect within each service user and tested it against change in both PGSI and Core-10.[9]

[9] For change in PGSI scores, the mixed model has an AIC (Akaike Information Criterion) of 83,882 and the standard linear regression model has an AIC of 48,862. This shows that the mixed model is a poorer fit to the data than the linear regression model. For change in Core-10 scores, the AIC is 87,618 for the mixed models and 52,609 for the linear regression model and there was no strong evidence that the random effects in this model were informative. For these reasons, we chose to use linear regression models with each referral treated as an independent observation.

We used multi-variable logistic regression to assess how variables were associated with reliable and clinically significant changes in Core-10 and PGSI (as defined in section 3.7), and no reliable change or reliable deterioration in both measures. All dependent variables are listed in Table 9.

***Table 9: Dependent variables for logistic regression analysis***

| Dependent variables |
|---|
| Reliable and clinically significant change in PGSI score using the PG >=8 threshold |
| Reliable and clinically significant change in PGSI score using a PGSI threshold of >=9 |
| Reliable change in PGSI score |
| No reliable change or reliable deterioration in PGSI score |
| Reliable and clinically significant change in Core-10 score |
| No reliable change or reliable deterioration in Core-10 score |
| Reliable and clinically significant change in both Core-10 and PGSI score |
| No reliable change or reliable deterioration in both Core-10 and PGSI score |
| Treatment completion (1 = completed. 0 = drop out or discontinued) |
| Brief referrals (<=3 attended appointments) |

The independent variables for the logistic regression are listed in Table 10; analysis was conducted separately for gamblers and affected others. Core-10 baselines were excluded from analysis of Core-10 change, and PGSI baselines were excluded from analysis of PGSI change.

***Table 10: Independent variables for logistic regression analysis***

| Category | Variable | Type |
|---|---|---|
| Referral | Number of attended appointments between first and last measure** | Scale |
| | Brief referrals (<=3 attended appointments) * | Categorical |
| | Total planned appointments missed** | Scale |
| | Treatment completion (1 = completed. 0 = drop out or discontinued)** | Categorical |
| service user | Known recurrence (whether service user had previously had treatment) | Categorical |
| | Gender | Categorical |
| | Age (age at referral) | Scale |
| | Referral number | Scale |
| Measures | PGSI behaviour domain (1,2,3,4) score at first appt | Scale |
| | PGSI recognition and guilt (5,9) score at first appt | Scale |
| | PGSI external consequences (6,7,8) score at first appt | Scale |
| | Core-10 functioning domain (2,3,4) score at first appt | Scale |
| | Core-10 depression and anxiety domain (1,5,7,8,9,10) score at first appt | Scale |
| | Core-10 risk-to-self item (6) score at first appt | Scale |

** excluded from brief referral model, ** excluded from treatment completion model*

To test whether differences existed for gamblers and affected others, we ran four logistic regression models as above, with client type added as a factor, and the following dependent variables:

- Treatment completion
- Brief referrals
- Clinically significant change in Core-10
- No reliable change or reliable deterioration in Core-10

### Appointment level analysis

### Reliability analysis

We calculated the reliability of responses to the items in the Core-10 and PGSI at the first attended appointment using Cronbach's alpha (a measure of internal consistency). This provided the information needed for the reliable change calculations, as reported in section 2.8.

We used a linear mixed model analysis to assess the impact of attended treatment appointments on change in PGSI and Core-10 scores and domains (dependent variables listed in Table 11) over time. This analysis was conducted separately for gamblers and affected others so that we could compare the PGSI and Core-10 models.

***Table 11: Dependent variables for linear mixed model***

| Measure | Variable |
|---|---|
| PGSI | Change in PGSI score between first and current appointment |
| | Change in PGSI behaviour domain score between first and current appointment |
| | Change in PGSI recognition and guilt score between first and current appointment |
| | Change in PGSI external consequences score between first and current appointment |
| Core-10 | Change in Core-10 score between first and current appointment |
| | Change in Core-10 functioning domain score between first and current appointment |
| | Change in Core-10 depression/anxiety domain score between first and current appointment |

Gender, age, treatment recurrence, completed treatment, attended appointment order, squared attended appointment order, previous appointment attendance, whether service user had multiple referrals included in the data and, if so, the order of these referrals (referral number) were all fixed effects (Table 12).

*Table 12: Fixed Effects for linear mixed model*

| Category | Variable | Type |
|---|---|---|
| Referral | Treatment completion (1 = completed. 0 = drop out or discontinue) | Categorical |
| Service user | Known recurrence (whether service user had previously had treatment) | Categorical |
| | Gender | Categorical |
| | Age (age at referral) | Scale |
| | Whether service user had multiple referrals | Categorical |
| | Referral number | Scale |
| Appointments | Attended appointment order | Scale |
| | Squared attended appointment order (quadratic) | Scale |
| | Previous appointment attendance | Categorical |

Attended appointment order was calculated using the attendance values of each appointment within a referral and ordering them according to date. For example, if a client had attended appointments weekly starting on 01/05/20, the appointment on 01/05/20 would have an attended appointment order of 1 (the first attended appointment), the appointment on 08/05/20 would have an attended appointment order of 2, and so on. Squared attended appointment order was included to test whether the impact of treatment was non-linear.

For appointments that occurred on the same day but had different purposes and measures, we treated assessments as having occurred before treatment, and treatment as occurring before reviews. As the attended appointment order increases, the sample decreases as more service users will have completed their treatment programme and left treatment.

Previous appointment attendance was calculated by checking the attendance of the appointment immediately preceding the current one, by date, in the referral. Attended appointments were recorded as 1 and missed appointments as 0.

An intercept, attended appointment order, squared attended appointment order, previous appointment attendance and the order of the service user's referrals were treated as random effects within each referral.

A full description of the linear mixed model is provided in Annex 3.

### Analysis of Tier 4 treatment data

Gordon Moody provided PGSI measures from the residential treatment programme for men with a gambling addiction between 2018 and 2021 linked to stage of treatment (i.e. start, end or follow up). The data provided by Gordon Moody contained information on 214 service users of residential care, although this number reduced to 90 after only those records which had measures for both start and end of treatment were included, and further down to 58 once only records with a treatment length of between 12-15 weeks and a starting PGSI score of >=8 were included. As the data from Gordon

Moody was different to the Tier 3 data, we wanted to create the most comparable data from Tier 3. To make a roughly comparable group from Tier 3 we applied the following rules:

- All referrals had to be 12 weeks or more in length and could not last longer than 14 weeks.
- Service user must be a male problem gambler, and be defined as a Problem Gambler (i.e. have a PGSI score of >=8) at treatment start.
- Service user must have only one referral in the DRF and must have completed treatment.
- Service user must have had at least 6 attended appointment sessions (this was for Tier 3 treatment only and was to ensure data included only service users receiving regular Tier 3 treatment).

This resulted in the inclusion of 58 service users from Gordon Moody and 511 from Tier 3 treatment.

We used t-tests to investigate differences in PGSI scores at the start of treatment between residential care and Tier 3 treatment service users.

We used linear regression with change in PGSI score between start and end of treatment as the dependent variable and treatment mode (Tier 3 vs residential care) as the independent variable, to compare how PGSI changed during treatment.

Finally, we used chi square tests to compare the proportion of residential care service users who showed clinically significant change in PGSI with the proportion in Tier 3 treatment.

# 3 Overview of referrals in analysis

Across the three-year analysis period the most referrals started in the year 2019/20 (40.0% of the total). The data equate to over half (53.1%) of all individuals referred over the same period (Table 13). The difference between qualifying referrals and individuals referred was mainly due to the qualifying criteria that excludes referrals to Gordon Moody, includes only referrals that also closed within the period and referrals with less than two attended appointments with measures. These three criteria alone accounted for the removal from analysis of 10,468 referrals.

***Table 13: Qualifying referrals by year***

| Year | Individuals referred* | Qualifying referrals | % of individuals referred that qualify |
|---|---|---|---|
| 2018/19 | 8,453 | 4,714 | 55.8% |
| 2019/20 | 9,726 | 5,784 | 59.5% |
| 2020/21 | 9,046 | 3,964 | 43.8% |
| **Total** | **27,225** | **14,462** | **53.1%** |

**GambleAware 2021, Table 33.*

## 3.1 Service user characteristics

### Client type and gender

The majority of service users seen over the analysis period were gamblers (12,348; 85.4%). The remaining group were affected others. As shown in Table 14, large differences were observed between gender and client type whereby there were more male (84.2%) gamblers and more female (84.6%) affected others[10].

***Table 14: Gender by client type***

| | Gamblers | | Affected others | | Total | |
|---|---|---|---|---|---|---|
| **Gender** | **N** | **%** | **N** | **%** | **N** | **%** |
| Male | 10,396 | 84.2% | 326 | 15.4% | 10,722 | 74.1% |
| Female | 1,952 | 15.8% | 1,788 | 84.6% | 3,740 | 25.9% |
| Total | **12,348** | | **2,114** | | **14,462** | |

*NB: excludes those with missing gender or gender categories with less than 10 service users*

[10] Chi squared test X2(1, 14,462) = 4452.36, p < .001.

## Age

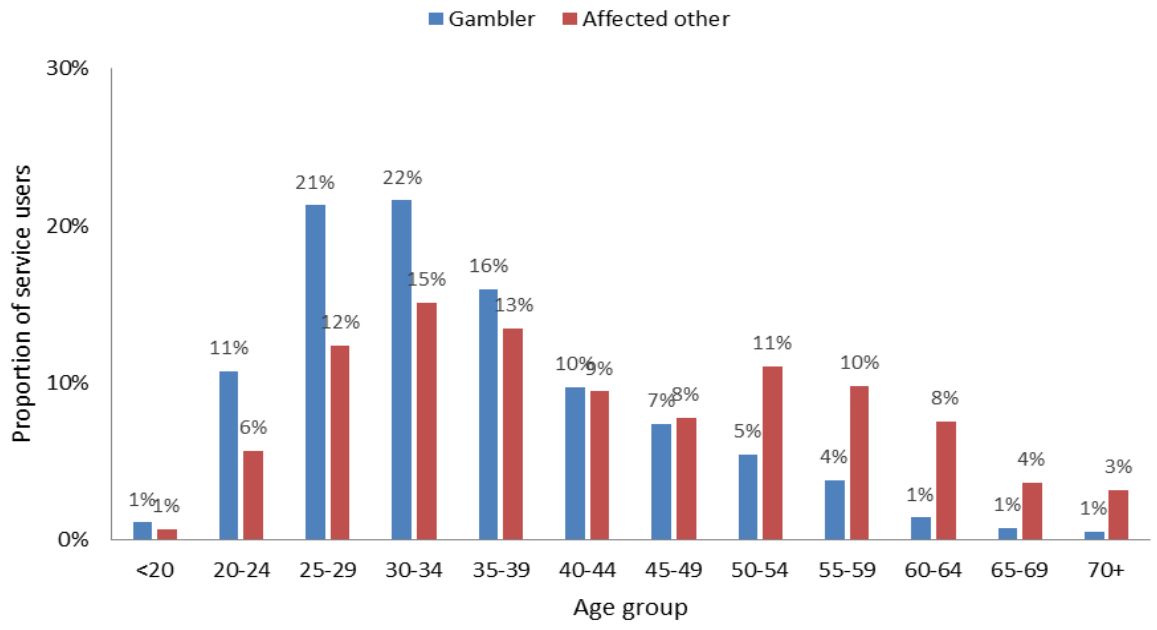
Service users had a mean age of 36.7 years (Standard Deviation (SD) = 11.5) at time of referral (Table 15). Gamblers had a mean age 7 years younger than affected others. Male gamblers (Mean (M) = 34.8, SD = 10.2) were younger than female gamblers (M = 39.8 SD = 11.8)[11], while male affected others (M = 45.9, SD = 15.6) were older than female affected others (M = 42.6, SD = 13.6)[12].

*Table 15: Gender by mean age*

| Gender | Gamblers | Affected others | Total |
|---|---|---|---|
| Male | 34.8 | 45.9 | 35.1 |
| Female | 39.8 | 42.6 | 41.2 |
| Total | **35.6** | **43.1** | **36.7** |

Most gamblers (54.8%) were aged under 35, with only 12.1% aged 50 or over. Affected others had an older profile with three times as many affected others aged 50 or over (35.4%), as shown in Figure 2. Gamblers (M = 35.6, SD = 10.6) were generally younger than affected others (M = 43.1, SD = 14.0)[13].

*Figure 2: Age group by client type*




[11] T test t(12346) = 19.661, p < .001

[12] T test t(2112) = -3.825, p < .001

[13] T test t(14460) = 28.7999, p < .001

## 3.2 Referral details

### Treatment recurrence

Nearly one quarter (24.1%) of all referrals were for recurring treatment, which is defined as service users self-reporting being previously seen by one of the providers within the NGTS (Table 16). More gamblers were re-entering treatment than affected others (25.6% compared to 15.2%)[14] .

*Table 16: Treatment recurrence by client type*

| | Gamblers | | Affected others | | Total | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| Recurring | 3,163 | 25.6% | 322 | 15.2% | 3,485 | 24.1% |
| First-time | 9,185 | 74.4% | 1,792 | 84.8% | 10,977 | 75.9% |
| Total | **12,348** | | **2,114** | | **14,462** | |

### Treatment completion

Of those service users who engaged in treatment (i.e. attended at least two appointments), the majority completed scheduled treatment (80.8%)[15]. A higher proportion of affected others completed treatment than gamblers (Table 17)[16] .

*Table 17: Treatment completion by client type*

| | Gamblers | | Affected others | | Total | |
|---|---|---|---|---|---|---|
| Treatment end reason | N | % | N | % | N | % |
| Completed scheduled treatment | 9,829 | 79.6% | 1,860 | 88.0% | 11,689 | 80.8% |
| Dropped out | 197 | 1.6% | 44 | 2.1% | 241 | 1.7% |
| Discharged by agreement | 2,283 | 18.5% | 206 | 9.7% | 2,489 | 17.2% |
| Not known | 39 | 0.3% | 4 | 0.2% | 43 | 0.3% |
| Grand Total | **12,348** | | **2,114** | | **14,462** | |

## 3.3 Length of treatment

Gamblers had greater mean treatment length (11.9 weeks, SD = 8.1) than affected others (10.9 weeks, SD = 7.7)[17]. 62.1% of all treatment referrals lasted 12 weeks or less (Table 18). Those who

[14] Chi squared test $X^2(1, 14,462) = 106.4$, $p < .001$

[15] It should be noted that this figure will be higher than typically reported for the NGTS as it excludes service users who drop out after only having one appointment as they are outside the scope of this project.

[16] Chi squared test $X^2(1, 14,462) = 81.88$, $p < .001$

[17] T test $t(2973) = -5.747$, $p < .001$

completed treatment had greater treatment length (M = 12.2 weeks, SD = 7.9 weeks) than those who dropped out (M = 10.1 weeks, SD = 8.5 weeks)[18]. A higher proportion of service users who dropped out of treatment (24.7%) had referrals lasting less than a month than those who completed treatment (13.6%).

*Table 18: Length of treatment by client type and treatment completion*

| Length of treatment | Client Type | | Treatment Completion | | Total |
|---|---|---|---|---|---|
| | Gamblers | Affected others | Completed | Dropped Out | |
| 0-4 weeks | 15.1% | 18.9% | 13.6% | 24.7% | 15.6% |
| 5-12 weeks | 46.3% | 47.7% | 45.8% | 51.0% | 46.5% |
| 13-20 weeks | 26.2% | 24.3% | 28.1% | 16.6% | 25.9% |
| 21+ weeks | 12.5% | 9.2% | 12.6% | 7.7% | 12.0% |

## Attended appointments

The mean number of attended treatment appointments in a referral was 6.4 for affected others (SD = 3.4) and 6.6 for gamblers (SD = 3.6)[19]. As shown in Table 19, those service users who had referrals lasting less than a month typically had around 3 attended appointments.

*Table 19: Mean attended appointments by length of treatment and treatment completion*

| Length of treatment | Treatment Completion | | Total |
|---|---|---|---|
| | Completed | Dropped Out | |
| 0-4 weeks | 2.9 appts | 2.4 appts | 2.8 appts |
| 5-12 weeks | 6.1 appts | 3.7 appts | 5.6 appts |
| 13-20 weeks | 9.0 appts | 6.0 appts | 8.7 appts |
| 21+ weeks | 11.0 appts | 7.6 appts | 10.7 appts |

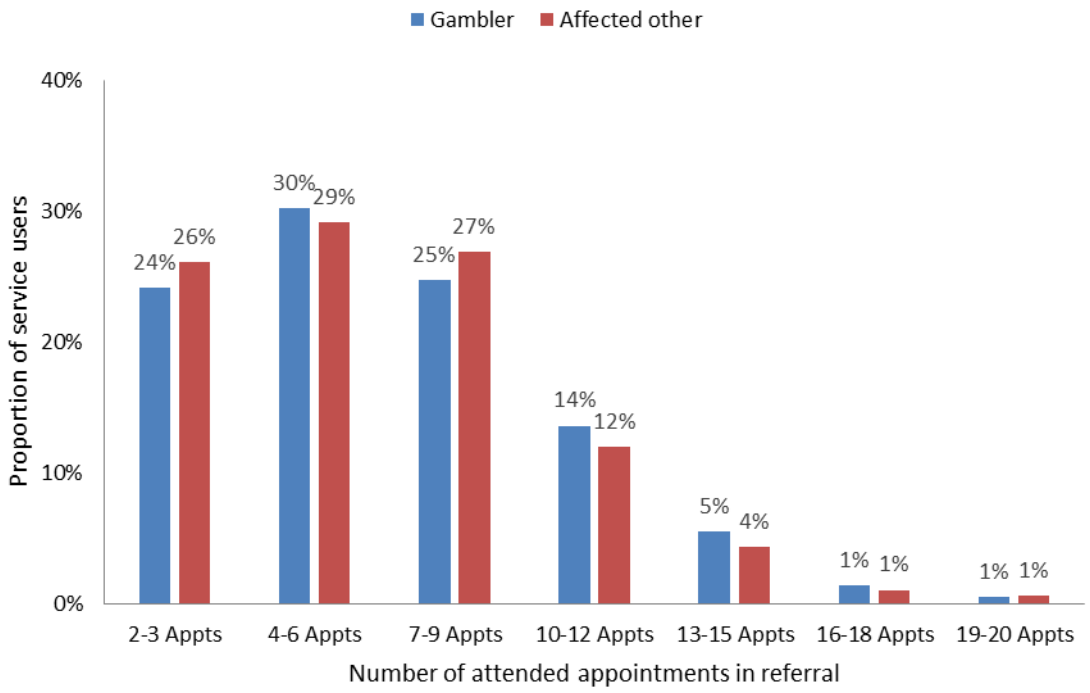
There was an even distribution of service users who had 2-3 appointments, 4-6 appointments, 7-9 appointments and 10 or more appointments, with each accounting for around one quarter of the total referrals (Figure 3).[20]

[18] T test t(3980) = -11.716, p < .001

[19] T test t(2954) = -2.995, p = .003

[20] NB: As detailed in section 2.4 a small number of very long referrals (>20 attended appointments) were removed from this analysis.

*Figure 3: Percentage of attended treatment appointments by client type*




### Missed and total appointments

Service users had a mean of around 8 planned appointments of which they attended around 6. Gamblers had slightly higher mean planned appointments (8.3, SD = 4.1) than affected others (7.5, SD = 3.9). Gamblers missed more appointments (M= 1.7, SD = 1.9) than affected others (M = 1.2, SD = 1.5), which equates to gamblers missing 20.8% of planned appointments compared to 15.8% for affected others.[21] There is little existing literature on attendance rates for psychotherapy but in a small sample of psychotherapy patients (N = 542), Defife et al (2010) found that 13% of planned appointments were not attended.

## 3.4 Severity at start of treatment

### Core-10

At the first attended appointment, 19.7% of service users were in the most severe category of Core-10, as shown in Table 20. Female gamblers had the highest proportion in the severe category (28.9%) with male affected others the lowest (5.8%). Overall, 81.6% of service users were in the clinical range, with female gamblers (90.3%) again the highest proportion and male affected others (54.6%) the lowest.

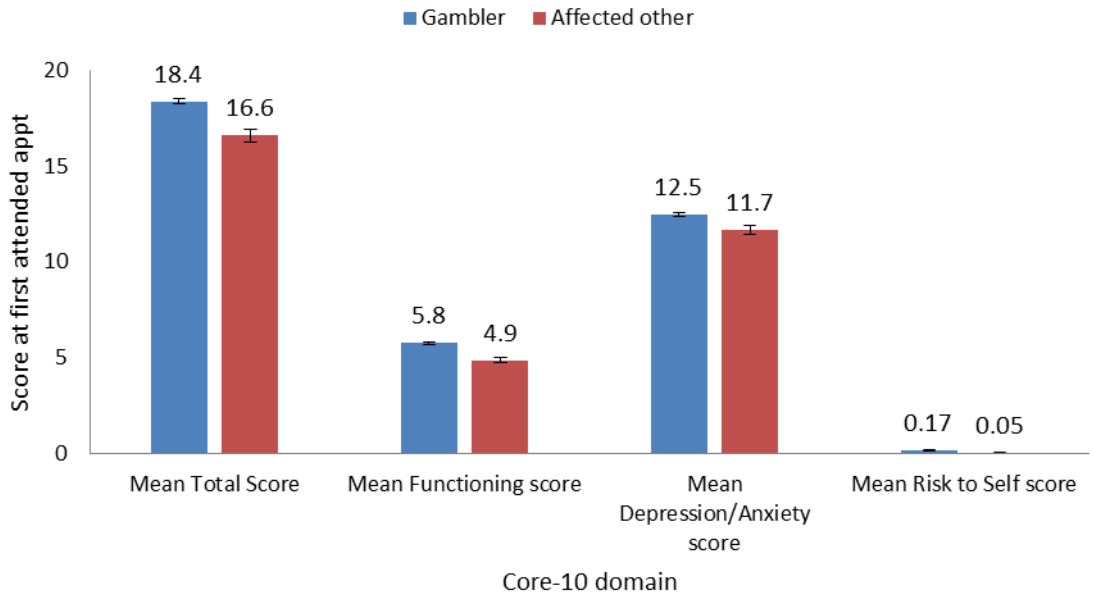
[21] T test t(14460) = -12.723, p < .001

*Table 20: Core-10 score categories at assessment by client type and gender*

| Core-10 Category | | Gambler | | Affected Other | | Total |
|---|---|---|---|---|---|---|
| | | Female | Male | Female | Male | |
| Non-clinical | Healthy | 2.5% | 7.1% | 4.7% | 32.8% | 6.8% |
| | Low-Level | 7.2% | 12.1% | 13.0% | 12.6% | 11.5% |
| Clinical | Mild | 15.2% | 19.7% | 22.4% | 16.6% | 19.4% |
| | Moderate | 22.0% | 22.5% | 25.1% | 17.2% | 22.6% |
| | Moderate to Severe | 24.2% | 19.1% | 20.9% | 15.0% | 19.9% |
| Severe | Severe | 28.9% | 19.5% | 13.8% | 5.8% | 19.7% |
| Mean Score | | 20.8 | 17.9 | 17.4 | 12.1 | 18.1 |
| N | | **1,952** | **10,396** | **1,788** | **326** | **14,462** |

On average, gamblers had higher Core-10 scores across each of the domains at the first attended appointment than affected others[22] (Figure 4). 97.3% of affected others responded 'not at all' to the risk to self question compared with 90.3% of gamblers.

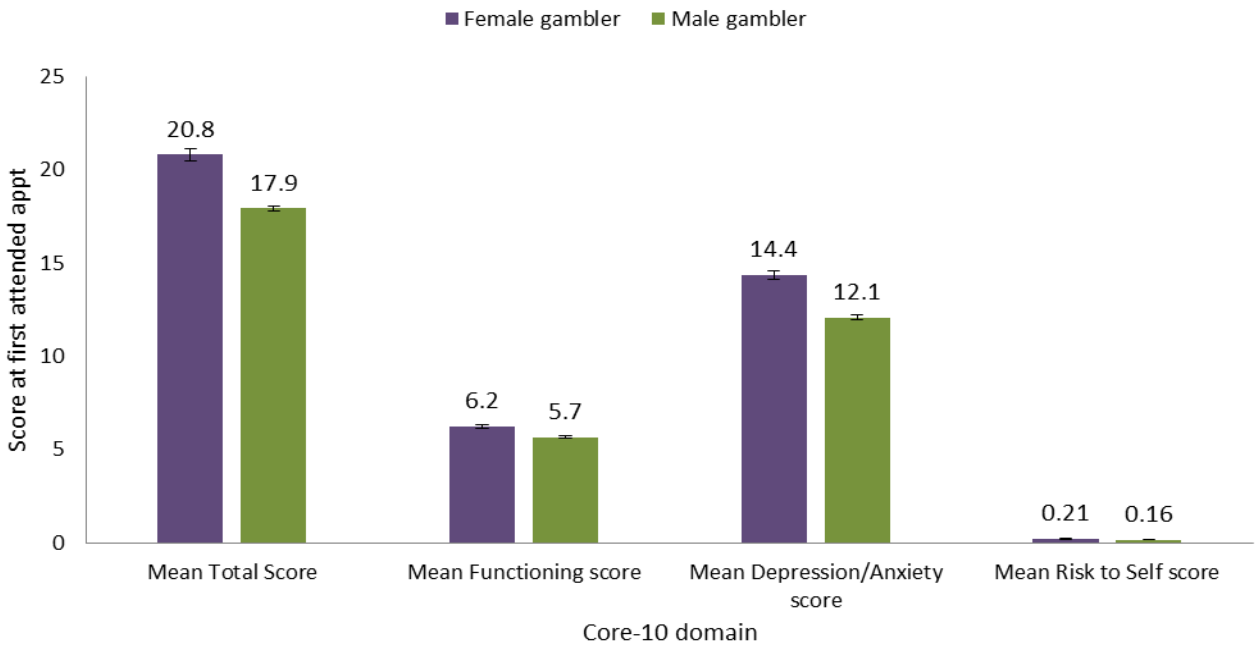
*Figure 4: Mean Core-10 score and domains (with 95% CI bars) at first attended appointment for gamblers and affected others*




Core-10 scores at first attended appointment for female gamblers were higher than for male gamblers, particularly for the depression/anxiety domain (Figure 5)[23]. 90.6% of male gamblers responded 'not at all' to the risk to self question compared with 88.7% of female gamblers.

[22] Functioning T test: t(14460) = -12.569, p < .001, Depression / Anxiety T test: t(14460) = -5.617, p < .001, Risk to self T test: t(14460) = -9.187, p < .001

*Figure 5: Mean Core-10 score and domains (with 95% CI bars) at first attended appointment for male and female gamblers*




## PGSI

94% of all gamblers were classed as problem gamblers by their PGSI score at the first attended appointment (Table 21). Female gamblers had slightly higher mean PGSI scores (19.5, SD = 5.8) at treatment start than male gamblers (18.9, SD = 6.3)[24].

*Table 21: PGSI score categories for gamblers at assessment by gender*

| PGSI score categories | Female | Male | Total |
|---|---|---|---|
| **non-problem gambler** | 0.6% | 0.6% | 0.6% |
| **low-risk gambler** | 0.4% | 0.8% | 0.8% |
| **moderate-risk gambler** | 2.9% | 4.4% | 4.2% |
| **problem gambler** | 96.2% | 94.1% | 94.4% |
| **Mean Score** | 19.5 | 18.9 | 19.0 |
| **N** | **1,952** | **10,396** | **12,348** |

Female gamblers had slightly higher scores in each of the three PGSI analysis domains (Figure 6)[25]. The means and standard deviations for PGSI and Core-10, by domain, for male and female gamblers are provided in Table 22

[23] Functioning T test: t(12346) = 7.474, p < .001, Depression / Anxiety T test: t(12346) = 15.818, p < .001, Risk to self T test: t(12346) = 3.586, p < .001

[24] T test t(12346) = 3.945, p < .001

*Figure 6: Mean PGSI score and domains (with 95% CI bars) at first attended appointment for male and female gamblers*

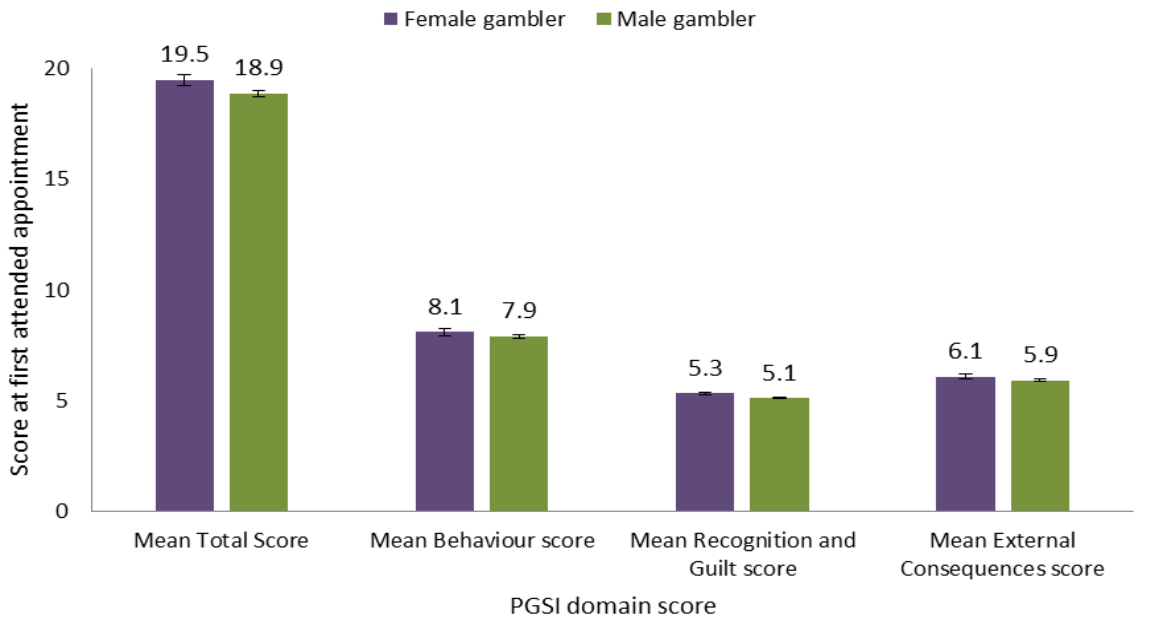



*Table 22: Mean PGSI and Core-10 domain scores at first attended appointment for gamblers by gender*

| | Maximum possible score | Female gamblers | | Male gamblers | |
|---|---|---|---|---|---|
| | | Mean | Std. Dev | Mean | Std. Dev |
| PGSI Score | 27 | 19.5 | 5.8 | 18.9 | 6.3 |
| PGSI Score (Behaviour) | 12 | 8.1 | 3.2 | 7.9 | 3.4 |
| PGSI Score (Recognition and Guilt) | 6 | 5.3 | 1.2 | 5.1 | 1.4 |
| PGSI Score (External Consequences) | 9 | 6.1 | 2.4 | 5.9 | 2.5 |
| Core-10 Score | 40 | 20.8 | 7.6 | 17.9 | 8.1 |
| Core-10 (Functioning) | 12 | 6.2 | 2.9 | 5.7 | 3.1 |
| Core-10 (Depression/Anxiety) | 24 | 14.4 | 5.6 | 12.1 | 5.9 |
| Core-10 (Risk to Self) | 4 | 0.2 | 0.7 | 0.2 | 0.6 |
| N | | 1,952 | | 10,396 | |

A Pearson correlation coefficient was computed to assess the linear relationship between PGSI and Core-10 score at first attended appointment. There was a moderate positive correlation between the two variables, indicating that clients/users with higher scores on the PGSI generally had higher scores on the Core-10[26].

[25] Behaviour T test: t(12346) = 2.469, p = .014, Recognition and Guilt T test: t(12346) = 6.737, p < .001, External consequences T test: t(12346) = 2.866, p = .004

[26] Pearson correlation r(12,346) = .46, p < .001.

# 4 Outcomes of Tier 3 Treatment for Gamblers

## 4.1 Change in gamblers' Core-10

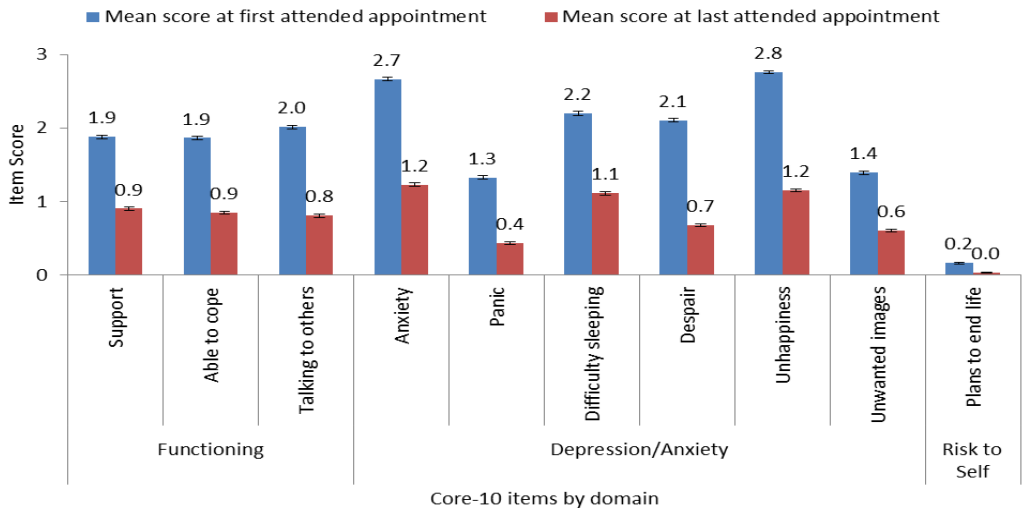
To assess the impact of treatment on gambler's Core-10 scores, we measured the difference (first minus last; positive values reflecting improvement while negative values reflect deterioration) in the Core-10 score at the first and last attended appointment. There were large reductions to the Core-10 total score and all domains between first and last attended appointment (Table 23). The mean difference between total Core-10 scores at last and first appointment was 10.6 (95% CI [10.4, 10.7]).

*Table 23: Pairwise t-test results comparing Core-10 scores at first and last attended appointment for Gamblers*

| | Paired difference between PGSI score at first and last appointment | | | |
|---|---|---|---|---|
| | | 95% CI of the Difference | | |
| | Mean difference | Lower | Upper | P-value |
| Core-10 Score | 10.56 | 10.41 | 10.72 | < .001 |
| Core-10 (Functioning) | 3.20 | 3.13 | 3.26 | < .001 |
| Core-10 (Depression and Anxiety) | 7.24 | 7.13 | 7.35 | < .001 |
| Core-10 (Risk to Self) | 0.13 | 0.12 | 0.14 | < .001 |

As illustrated in Figure 7, 'Anxiety' and 'Unhappiness' have the highest individual item scores at the first and last attended appointment, but still showed large reductions. Change was broadly consistent for all items, with each reducing by between 50-75% of the score at first attended appointment.

*Figure 7: Mean Core-10 item score (with 95% CI bars) at last appointment and change from first*[27]




[27] NB: items are sorted into domains and do not correspond in order to the standard item numbering of the Core-10.

We performed multi-variable linear regression with the difference in Core-10 score in the first and last available appointment (first minus last) as the dependent variable, meaning a positive change value would show the gamblers Core-10 score reducing between first and last appointment, as detailed in section 2.10 (Table 24).

The coefficient value represents the mean change in the dependent variable given a one unit change in the predictor variable. The effect size is measured using Partial Eta Squared, which is the proportion of variance in the outcome associated with the estimated effect.

***Table 24: Regression output for the difference in Core-10 scores at first and last appointment for gamblers***

| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 11.04 | < .001 | 0.049 | 10.18 | 11.90 |
| Total attended appointments | 0.23 | < .001 | 0.006 | 0.18 | 0.29 |
| Brief referral (<=3 attended appts) | -3.96 | < .001 | 0.023 | -4.42 | -3.51 |
| Total missed appointments | -0.45 | < .001 | 0.009 | -0.54 | -0.37 |
| Completed Treatment Status (completed) | 1.32 | < .001 | 0.003 | 0.91 | 1.74 |
| Known treatment recurrence | -0.85 | < .001 | 0.002 | -1.23 | -0.47 |
| Referral number | -1.10 | < .001 | 0.002 | -1.53 | -0.68 |
| Gender (Female) | 0.10 | .632 | <.001 | -0.31 | 0.52 |
| Service user age at referral | 0.00 | .600 | <.001 | -0.01 | 0.02 |
| Adjusted $R^2$ | 0.100 | | | | |

*Dependent variable (difference in Core-10 scores at first and last appointment), N = 12,438.*

The factor with the largest effect on difference in Core-10 was whether the gambler had a brief referral[28], with those who had brief referrals estimated to have almost a 4-point smaller difference in Core-10 than those who had longer referrals.

Attended appointments and completed treatment referrals were both estimated to increase Core-10 difference (i.e. have a positive impact). Gamblers in subsequent treatment (both in terms of treatment recurrence and higher referral number) and those who missed appointments were estimated to have smaller reductions in Core-10 by treatment end. There was no strong evidence that age or gender had large effects on Core-10 difference.

The model explains 5.3% of observed variance when only the attended appointments are included in the model. This increases to 10.0% when all variables are included; meaning that 10% of the variation in difference in Core-10 is explained by the variables included in the model. It is hard to interpret whether this is lower than we might expect, given the range of factors that might impact someone's mental health, but it does suggest that there are many external factors that may determine an individual's response to treatment.

[28] VIF = 1.750 so no issue of multi-collinearity with total attended appointments

## 4.2 Change in gamblers' Core-10 domains

We used multi-variable regression models, using the same set of control variables as for the total Core-10 score, to measure what associates with differences in Core-10 Functioning and Core-10 Depression/Anxiety scores between first and last attended appointment.

All variables had the same effect direction as those for difference in Core-10 total score but with much smaller effect sizes, as would be expected as total Core-10 has a greater distribution of possible scores. For Core-10 Functioning, the model explains 6.7% of observed variance.

***Table 25: Regression output for the difference in Core-10 Functioning Scores at first and last appointment***

| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 3.09 | < .001 | 0.024 | 2.74 | 3.43 |
| Total attended appointments | 0.08 | < .001 | 0.004 | 0.06 | 0.11 |
| Brief referral (<=3 attended appts) | -1.10 | < .001 | 0.011 | -1.28 | -0.91 |
| Total missed appointments | -0.18 | < .001 | 0.008 | -0.21 | -0.14 |
| Completed Treatment Status (completed) | 0.49 | < .001 | 0.003 | 0.32 | 0.66 |
| Known treatment recurrence | -0.36 | < .001 | 0.002 | -0.51 | -0.21 |
| Referral number | -0.29 | .001 | 0.001 | -0.46 | -0.12 |
| Gender (Female) | -0.05 | .583 | <.001 | -0.21 | 0.12 |
| Service user age at referral | 0.01 | .092 | <.001 | 0.00 | 0.01 |
| Adjusted $R^2$ | 0.067 | | | | |

*Dependent variable (difference in Core-10 Functioning scores at first and last appointment), N = 12,438.*

For Core-10 Depression/Anxiety, the model explains 8.8% of observed variance. Again, the variables had the same effect directions with small effect sizes to the model for difference in Core-10 total score.

***Table 26: Regression output for the difference in Core-10 Depression/Anxiety Scores at first and last appointment***

| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 7.84 | < .001 | 0.048 | 7.23 | 8.46 |
| Total attended appointments | 0.14 | < .001 | 0.004 | 0.10 | 0.18 |
| Brief referral (<=3 attended appts) | -2.84 | < .001 | 0.023 | -3.17 | -2.52 |
| Total missed appointments | -0.29 | < .001 | 0.007 | -0.35 | -0.22 |
| Completed Treatment Status (completed) | 0.87 | < .001 | 0.003 | 0.58 | 1.17 |
| Known treatment recurrence | -0.47 | .001 | 0.001 | -0.74 | -0.19 |
| Referral number | -0.79 | < .001 | 0.002 | -1.09 | -0.48 |
| Gender (Female) | 0.14 | .376 | <.001 | -0.16 | 0.43 |
| Service user age at referral | 0.00 | .783 | <.001 | -0.01 | 0.01 |
| Adjusted $R^2$ | 0.088 | | | | |

*Dependent variable (difference in Core-10 Depression/Anxiety scores at first and last appointment), N = 12,438*

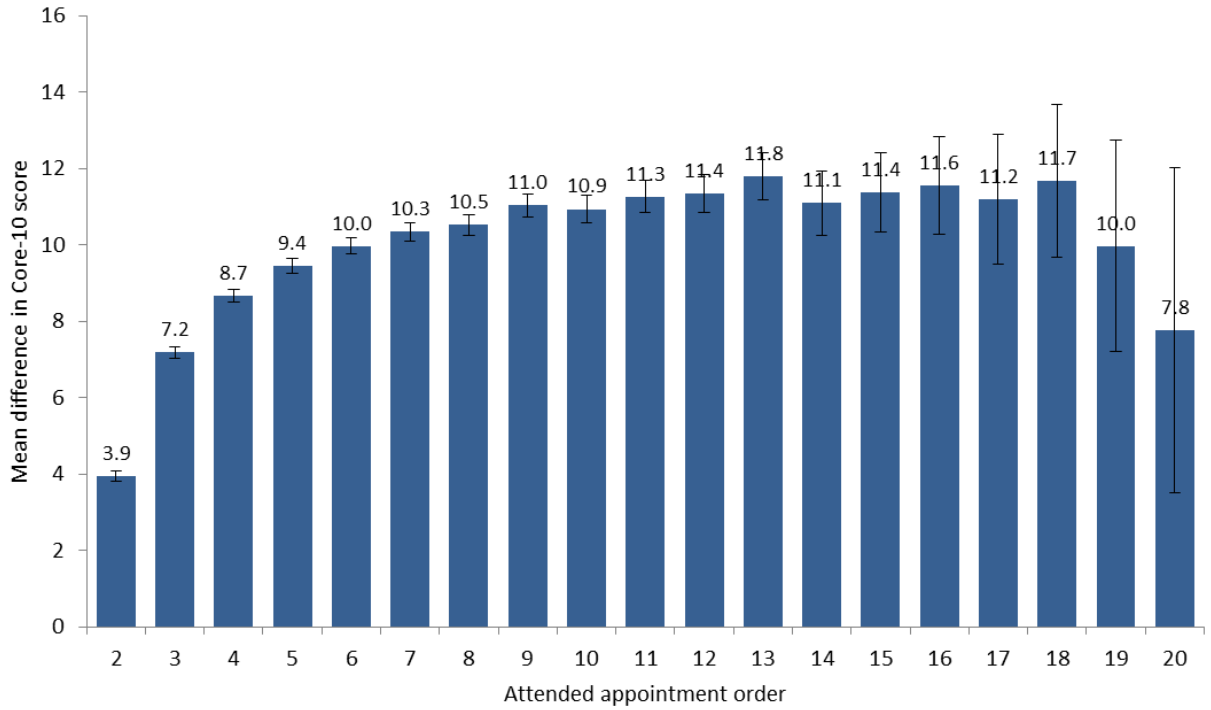
## 4.3 Gamblers' Core-10 across all appointments

This section uses the appointment level data to assess how treatment impacts Core-10 scores at each attended appointment. The variables used here differ to previous sections as they are no longer aggregates of a referral (first Core-10 score, total attended appointments, etc.) but instead are based on the data from each individual appointment for each gambler.

We calculated difference in Core-10 at each appointment by subtracting the initial baseline score (recorded at the first attended appointment) from the score at each attended appointment. So this is the score at each appointment compared to the score at the first appointment; not the score at each appointment compared to the score at the previous appointment. This gives us a figure for Core-10 difference at each appointment.

The greatest difference in Core-10 score compared to the first appointment was at appointment 13 (M = 11.8, 95% CI [11.2,12.4])). After appointment 13, the mean difference in Core-10 score starts to decrease (meaning the score is closer to the baseline score at first appointment) with the lowest difference observed at appointment 20 (M = 7.8, 95% CI [3.5,12.0]). However, there is considerable uncertainty in the estimates for later appointments which preclude drawing firm conclusions on changes in scores (Figure 8).

***Figure 8: Mean difference in Core-10 score (with 95% CI bars) by attended appointment order for gamblers***




We tested the apparent non-linear impact of treatment using a mixed model with the difference in Core-10 at each appointment compared to the Core-10 at the first appointment as the dependent variable and including both the attended appointment order and the squared attended appointment

order (to account for non-linearity) as variables. The parameter estimates for change in Core-10 score are provided in Table 27.

The positive coefficient for the linear term (attended appointment order) and the negative coefficient for the quadratic term (squared attended appointment order) indicate that difference in Core-10 score increases with initial attended appointments before stabilising with further attended appointments.

The maximum difference in Core-10 score is attained at (keeping all other covariates fixed) $\frac{1.959}{2\times.069} =$ 14.1 attended appointments, rounded to 14.

***Table 27: Parameter Estimates of linear mixed model with quadratic effect on difference in Core-10 score for gamblers***

| Parameter | Estimate | Std. Error | t | p value | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | 1.727 | 0.325 | 5.310 | <.001 | 1.090 | 2.365 |
| Attended appointment order | 1.959 | 0.025 | 77.794 | <.001 | 1.909 | 2.008 |
| Squared attended appointment order | -0.069 | 0.002 | -37.365 | <.001 | -0.073 | -0.066 |
| Previous appointment attendance | -0.017 | 0.057 | -0.296 | .768 | -0.129 | 0.095 |
| Completed treatment | 0.727 | 0.154 | 4.733 | <.001 | 0.426 | 1.028 |
| Gender (Female) | -0.144 | 0.198 | -0.728 | .467 | -0.531 | 0.244 |
| Service user age at referral | 0.003 | 0.007 | 0.435 | .663 | -0.010 | 0.016 |
| Known treatment recurrence | -0.761 | 0.139 | -5.488 | <.001 | -1.033 | -0.489 |
| Service user with multiple referrals | -0.122 | 0.265 | -0.461 | .645 | -0.642 | 0.398 |
| Referral number (by service user) | -0.845 | 0.176 | -4.803 | <.001 | -1.190 | -0.500 |

*Dependent variable (difference in Core-10 score), N = 69,183*

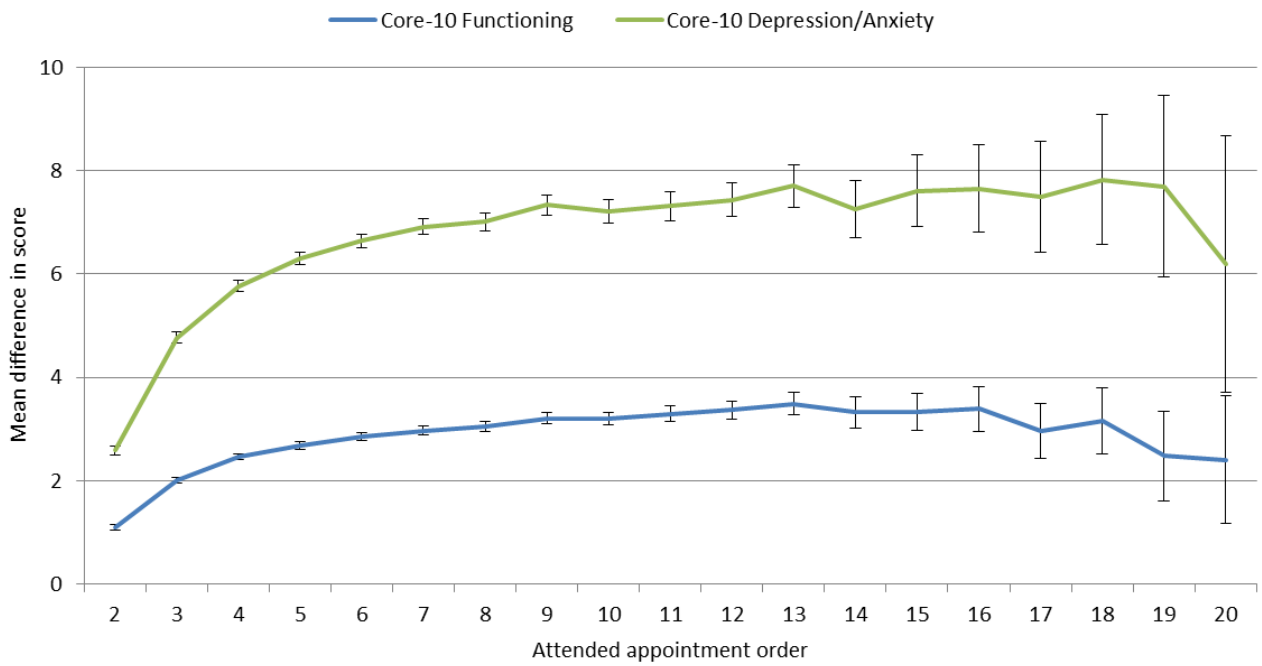
In this model, the parameter estimates from the linear mixed models (Table 27) show the variable with the largest effect size was attended appointment order, followed by squared attended appointment order. Interestingly, there was no strong evidence that whether the previous appointment was missed or attended associated with change in Core-10. This suggests that individual missed appointments may not impact how a gambler responds at the following appointment in terms of their Core-10 scores.

Treatment recurrence (those in repeated treatment) associated with smaller differences in Core-10 score, while treatment completion associated with larger differences in Core-10 (so gamblers who complete treatment show larger reductions in Core-10). There was no strong evidence that gender and age of the gambler associated with change in Core-10.

Both Core-10 Functioning and Core-10 Depression/Anxiety domain scores appear to associate with attended treatment in very similar ways (Figure 9), an increase in difference at the beginning of

treatment, followed by a plateau and then a gradual decrease in difference at later appointments (though with greater estimation uncertainty).

***Figure 9: Mean difference in Core-10 domain scores (with 95% CI bars) by attended appointment order for gamblers***




The maximum difference in Core-10 Functioning score is attained at $\frac{0.579}{2\times0.020} = 14.3$ appointments (Table 28), while for Core-10 Depression/Anxiety score it is attained at $\frac{1.350}{2\times.049} = 13.9$ appointments (Table 29).

***Table 28: Parameter Estimates of linear mixed model with quadratic effect on difference in Core-10 functioning score for gamblers***

| Parameter | Estimate | Std. Error | t | p value | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | 0.194 | 0.135 | 1.432 | .152 | -0.071 | 0.459 |
| Attended appointment order | 0.579 | 0.010 | 59.929 | <.001 | 0.561 | 0.598 |
| Squared attended appointment order | -0.020 | 0.001 | -28.816 | <.001 | -0.022 | -0.019 |
| Previous session attendance | 0.020 | 0.023 | 0.888 | .374 | -0.024 | 0.064 |
| Completed Treatment | 0.339 | 0.064 | 5.342 | <.001 | 0.215 | 0.464 |
| Gender (Female) | -0.049 | 0.082 | -0.596 | .551 | -0.211 | 0.112 |
| Client age at referral | 0.005 | 0.003 | 1.591 | .112 | -0.001 | 0.010 |
| Known treatment recurrence | -0.301 | 0.057 | -5.255 | <.001 | -0.413 | -0.189 |
| Client with multiple episodes | -0.166 | 0.111 | -1.494 | .135 | -0.384 | 0.052 |
| Referral number (by client) | -0.147 | 0.074 | -1.999 | .046 | -0.291 | -0.003 |

*Dependent variable (difference in Core-10 Functioning score), N = 69,183*

Each subsequent referral, as identified by referral number which increases for each subsequent referral a gambler has within the data, have slightly lower difference in Core-10 Depression/Anxiety. This suggests that treatment may not be as effective at helping reduce the anxiety of gamblers for those who have returned to treatment. Effects of all other variables on the scores are similar to the difference in total Core-10 score model.

***Table 29: Parameter Estimates of linear mixed model with quadratic effect on difference in Core-10 depression/anxiety score for gamblers***

| Parameter | Estimate | Std. Error | t | p value | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | 1.511 | 0.233 | 6.474 | <.001 | 1.054 | 1.968 |
| Attended appointment order | 1.350 | 0.018 | 73.615 | <.001 | 1.314 | 1.385 |
| Squared attended appointment order | -0.049 | 0.001 | -36.160 | <.001 | -0.051 | -0.046 |
| Previous session attendance | -0.032 | 0.042 | -0.764 | .445 | -0.114 | 0.050 |
| Completed Treatment | 0.423 | 0.110 | 3.837 | <.001 | 0.207 | 0.640 |
| Gender (Female) | -0.105 | 0.141 | -0.741 | .458 | -0.381 | 0.172 |
| Client age at referral | -0.002 | 0.005 | -0.427 | .669 | -0.012 | 0.007 |
| Known treatment recurrence | -0.473 | 0.100 | -4.737 | <.001 | -0.668 | -0.277 |
| Client with multiple episodes | 0.078 | 0.189 | 0.410 | .682 | -0.293 | 0.449 |
| Referral number (by client) | -0.696 | 0.127 | -5.490 | <.001 | -0.945 | -0.448 |

*Dependent variable (difference in Core-10 Depression/Anxiety score), N = 69,183*

## 4.4 Reliable and clinically significant change in Core-10

In section 2.8, reliable change and clinically significant change for Core-10 was defined. At treatment start, 85% of gamblers (10,513) met the clinical threshold of a Core-10 score of equal to or greater than 11.

***Table 30: Rates of reliable and clinically significant change in Core-10 for gamblers***

| | Gamblers above Clinical threshold | |
|---|---|---|
| | N | % |
| Reliable and clinically significant improvement | 5,907 | 56.2% |
| Reliable improvement only | 1,840 | 17.5% |
| No reliable change | 2,608 | 24.8% |
| Reliable deterioration | 158 | 1.5% |
| Total | **10,513** | **100.0%** |

56% of all gamblers who were above the clinical threshold at treatment start showed reliable and clinically significant change in Core-10 at treatment end (change of >=7 and end score of <11).

Table 31 shows the outputs from logistic regression with clinically significant Core-10 change in gamblers as the dependent variable.

***Table 31: Logistic regression output for clinically significant Core-10 change in gamblers***

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Completed Treatment Status (completed) | 1.67 | < .001 | 1.49 | 1.87 |
| Total missed appointments | 0.87 | < .001 | 0.85 | 0.89 |
| Total attended appointments | 1.05 | < .001 | 1.03 | 1.06 |
| Brief referral (<=3 attended appts) | 0.34 | < .001 | 0.30 | 0.39 |
| Treatment recurrence | 0.70 | < .001 | 0.63 | 0.77 |
| Referral number | 0.82 | .001 | 0.73 | 0.92 |
| Service user age at referral | 0.99 | < .001 | 0.99 | 1.00 |
| Gender (Female) | 0.66 | < .001 | 0.59 | 0.73 |
| Baseline PGSI Score (Behaviour) | 0.99 | .131 | 0.97 | 1.00 |
| Baseline PGSI Score (Recognition and Guilt) | 1.09 | < .001 | 1.04 | 1.14 |
| Baseline PGSI Score (External Consequences) | 0.94 | < .001 | 0.92 | 0.97 |
| Constant | 2.11 | < .001 | | |

*Dependent variable (clinically significant Core-10 change), N (gamblers clinical Core-10 at start of treatment) = 10,513*

Effects sizes were largest for brief referrals. Brief referrals reduced the odds of gamblers showing clinically significant change in Core-10 compared to longer referrals by 66%. Gamblers who complete treatment have higher odds of showing clinically significant change in Core-10 compared to those who drop out by 67%.

The odds of clinically significant change in Core-10 were lower for female compared to male gamblers, older compared to younger gamblers and those with treatment recurrence compared to those in first time treatment. For the difference by gender, this may be in part due to higher mean starting scores that female gamblers have compared to male gamblers, meaning they would need to show greater reductions to get below the clinical threshold.

The baseline scores for PGSI recognition and guilt were positively associated with a clinically significant change in Core-10. This means that gamblers with higher scores for ‘recognition of problem’ and ‘feelings of guilt’ at treatment start had higher odds of showing clinically significant change in Core-10 at the end of treatment. There was no strong evidence that the PGSI behaviour domain associated with clinically significant change in Core-10, but gamblers with higher PGSI external consequences had lower odds of showing clinically significant improvement in Core-10.

## 4.5 Change in gamblers’ PGSI

To assess the impact of treatment on gambler’s PGSI scores we measured the difference (first minus last) in the PGSI score at the first attended appointment (typically an assessment) and at the last attended appointment at which a measure was available. There were large reductions in the

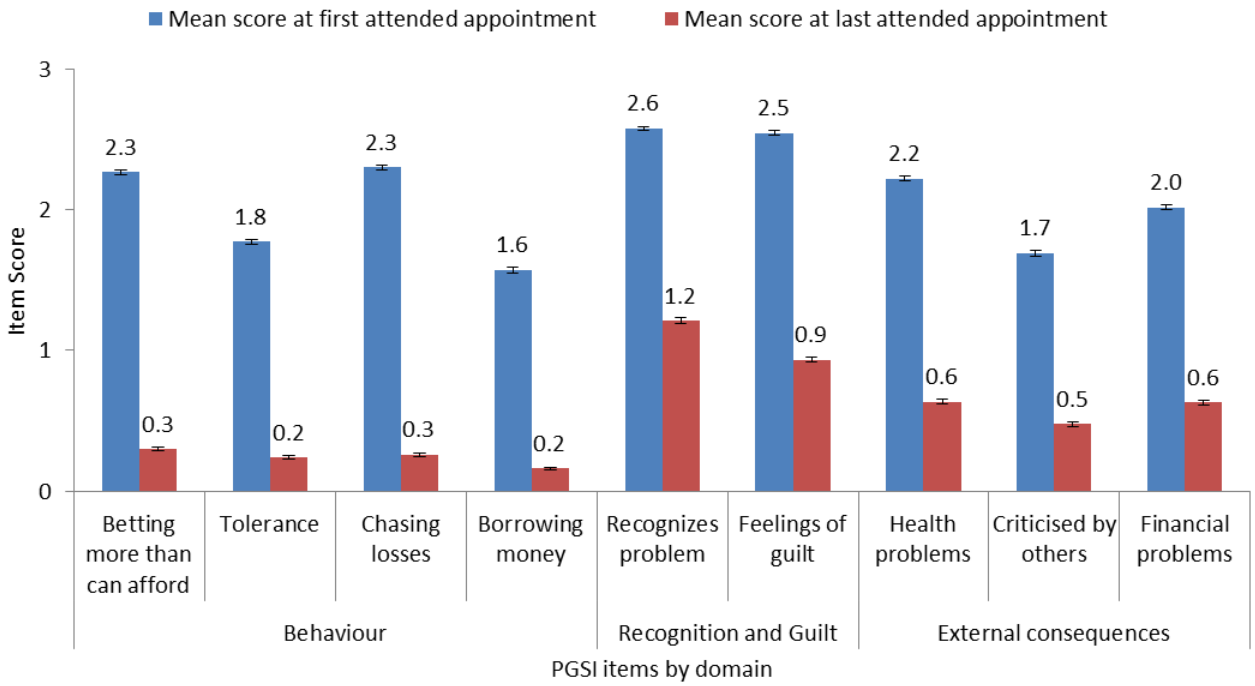
PGSI total score and all domains between first and last attended appointment (Table 32). The mean difference between total PGSI score at first and last appointment was 14.1 (95% CI [14.0, 14.3]).

***Table 32: Pairwise t-test results comparing PGSI scores at first and last attended appointment***

| | Paired difference between PGSI score at first and last appointment | | | |
|---|---|---|---|---|
| | | 95% CI of the Difference | | |
| | Mean difference | Lower | Upper | P-value |
| PGSI Score | 14.12 | 13.99 | 14.26 | < .001 |
| PGSI Score (Behavioural) | 6.96 | 6.89 | 7.03 | < .001 |
| PGSI Score (All Consequence) | 7.18 | 7.09 | 7.26 | < .001 |
| PGSI Score (Recognition and Guilt) | 2.98 | 2.94 | 3.02 | < .001 |
| PGSI Score (Other Consequences) | 4.20 | 4.14 | 4.25 | < .001 |

As illustrated in Figure 10, the PGSI items for '*Recognises problem*' and *'Feelings of guilt'* have the highest scores at the first attended appointment and reduce the least proportionately between the first and last attended appointment.

***Figure 10: Mean PGSI item score (with 95% CI bars) at first and last attended appointment by domain[29]***




[29] NB: items are sorted into domains and do not correspond to the order in the standard item numbering of the PGSI.

*'Betting more than can afford'* and '*Chasing losses'* are the next highest scores at first appointment but show the largest reduction by last appointment. All the behaviour items are lower at last attended appointment and show proportionately higher reduction than any consequence item.

We performed multi-variable linear regression with the difference in PGSI score between the first and last available appointment (first minus last) as the dependent variable, meaning a positive change value would show the PGSI score reducing between first and last appointment, as detailed in section 2.10 (Table 33).

***Table 33: Regression output for the difference in PGSI Scores at first and last appointment***

| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 17.47 | < .001 | 0.149 | 16.73 | 18.21 |
| Total attended appointments | 0.16 | < .001 | 0.004 | 0.11 | 0.21 |
| Brief referral (<=3 attended appts) | -4.38 | < .001 | 0.038 | -4.77 | -3.99 |
| Total missed appointments | -0.23 | < .001 | 0.003 | -0.30 | -0.16 |
| Completed Treatment Status (completed) | 1.25 | < .001 | 0.004 | 0.90 | 1.61 |
| Known treatment recurrence | -0.69 | < .001 | 0.001 | -1.01 | -0.36 |
| Referral number | -1.73 | < .001 | 0.007 | -2.09 | -1.36 |
| Gender (Female) | -0.05 | .781 | <.001 | -0.41 | 0.31 |
| Service user age at referral | -0.05 | < .001 | 0.005 | -0.06 | -0.04 |
| Adjusted $R^2$ | 0.119 | | | | |

*Dependent variable (difference in PGSI scores at first and last appointment), N = 12,348*

The factor with the largest effect on difference in PGSI was whether the gambler had a brief referral[30], with those who had brief referrals estimated to have on average a 4.38 smaller difference in PGSI than those who had longer referrals.

Attended appointments and completed treatment referrals were both estimated to increase PGSI difference over time (i.e. have a positive impact). Gamblers in subsequent treatment (both in terms of treatment recurrence and higher referral number) and those who missed appointments were estimated to have smaller reductions in PGSI by treatment end. Older gamblers were estimated to have smaller differences in PGSI, while there were no clear differences by gender.

The model explains 11.9% of observed variance in the difference scores, meaning 11.9% of the variation in difference in PGSI between first and last session can be statistically explained by the factors included in the model.

[30] VIF = 1.750 so no issue of multi-collinearity with total attended appointments

## 4.6 Change in gamblers' PGSI domains

We also estimated the impact of treatment on difference in PGSI behaviour and PGSI Consequence (which includes both PGSI External Consequence and PGSI Recognition and Guilt) scores between first and last attended appointment (first minus last).

For differences in PGSI behaviour, the model explained only 5.8% of observed variance. All effects had the same direction as those for difference in PGSI total score but with much smaller effect sizes.

***Table 34: Regression output for the difference in PGSI Behaviour scores at first and last appointment***

| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 8.91 | < .001 | 0.143 | 8.52 | 9.29 |
| Total attended appointments | 0.05 | < .001 | 0.001 | 0.03 | 0.08 |
| Brief referral (<=3 attended appts) | -1.51 | < .001 | 0.017 | -1.72 | -1.31 |
| Total missed appointments | -0.04 | .037 | <.001 | -0.08 | 0.00 |
| Completed Treatment Status (completed) | 0.30 | .001 | 0.001 | 0.12 | 0.49 |
| Known treatment recurrence | -0.29 | .001 | 0.001 | -0.46 | -0.12 |
| Referral number | -0.90 | < .001 | 0.007 | -1.10 | -0.71 |
| Gender (Female) | 0.05 | .616 | <.001 | -0.14 | 0.23 |
| Service user age at referral | -0.03 | < .001 | 0.006 | -0.04 | -0.02 |
| Adjusted $\boldsymbol{R^2}$ | 0.058 | | | | |

*Dependent variable (difference in PGSI Behaviour scores at first and last appointment), N = 12,348*

For changes to PGSI Consequence, the model explained 13.0% of observed variance. The higher $R^2$ value is in part caused by the larger effect sizes of brief referrals, attended and missed appointments and treatment completion on difference in PGSI consequence scores compared to difference in PGSI behaviour scores. This suggests that continued treatment may have a greater impact on PGSI consequence scores than it does on PGSI behaviour scores.

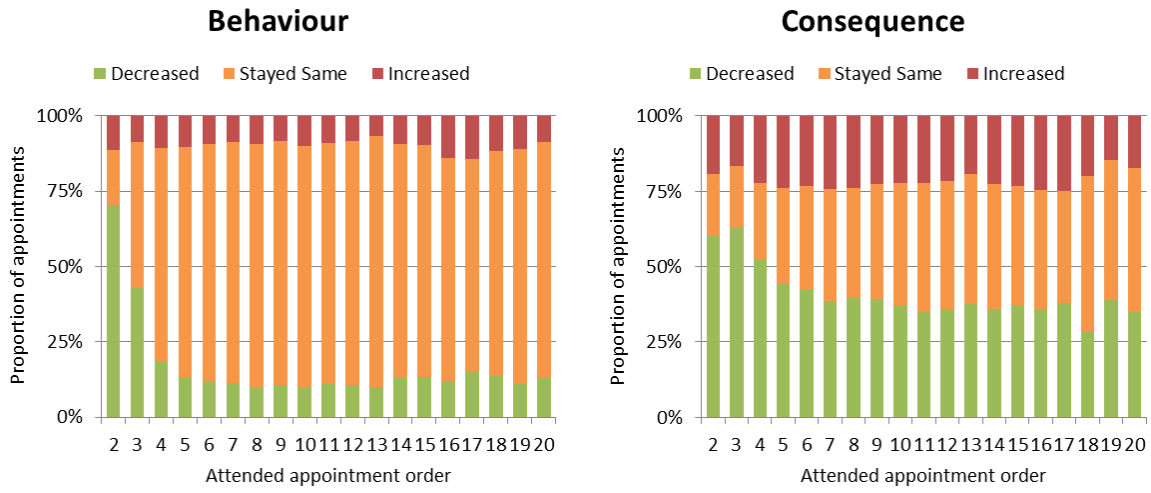
***Table 35: Regression output for the difference in PGSI Consequence scores at first and last appointment***

| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 8.56 | < .001 | 0.096 | 8.10 | 9.03 |
| Total attended appointments | 0.11 | < .001 | 0.004 | 0.08 | 0.14 |
| Brief referral (<=3 attended appts) | -2.87 | < .001 | 0.041 | -3.11 | -2.62 |
| Total missed appointments | -0.19 | < .001 | 0.005 | -0.23 | -0.14 |
| Completed Treatment Status (completed) | 0.95 | < .001 | 0.006 | 0.73 | 1.17 |
| Known treatment recurrence | -0.39 | < .001 | 0.001 | -0.60 | -0.19 |
| Referral number | -0.82 | < .001 | 0.004 | -1.05 | -0.59 |
| Gender (Female) | -0.10 | .391 | <.001 | -0.32 | 0.13 |
| Service user age at referral | -0.02 | < .001 | 0.003 | -0.03 | -0.02 |
| Adjusted $\boldsymbol{R^2}$ | 0.130 | | | | |

*Dependent variable (difference in PGSI Consequence scores at first and last appointment), N = 12,348*

Figure 11 shows the direction of change for PGSI behaviour and PGSI consequence domains at each attended session compared to the previous session. Behaviour domain items show a strong decrease in the first three appointments before staying consistently low throughout treatment, with the majority staying the same between appointments. Conseqence domain items show a much slower change at the start of treatment, with higher rates of both decrease and increase from previous session than the behaviour items, suggesting much greater variability across the entire treatment journey.[31]

***Figure 11: Changes in score compared to previous appointment by PGSI domain***




At the first attended appointment, the PGSI works as a valid screening tool and categorises the vast majority of gamblers who have gambling problems as problem gamblers.

As gamblers go through treatment, however, the PGSI does not function as it does in other contexts. Behavioural items are quickly reduced and remain low throughout treatment, while consequence items are far more variable and (particularly those covering 'recognition of problem' and 'feelings of guilt') remain much higher throughout treatment. This two-part construct is similar to that found in other gambling measures with one construct focused on loss of control and dependency and one focused on the adverse consequences resulting from dependency (Abbott and Volberg, 2006).

As a screening tool, the PGSI typically asks participants to self-assess their gambling behaviour over the past 12 months by scoring themselves against nine questions. Samuelson et al (2019)

[31] NB: the longer treatment lasts, the smaller the sample size for the outcomes data (as most people do not have referrals lasting more than 7 attended appointments). The greater variability seen in later appointments could be caused by smaller sample sizes.

identified the possibility of ambiguity in responses, particularly in relation to item 9 ('feelings of guilt'). When used regularly with people with gambling problems in treatment, the potential for ambiguity increases. Locke et al (2013) identified that US students with problem gambling have "excessive interpersonal guilt". Treatment for problem gambling does not undo all the harms caused by a person's problem gambling and many will still have feelings of guilt, which may actually be a key motivational tool within treatment itself. So while the scores for gambling behaviours show a greater decrease during treatment, it would appear that treatment itself explains a greater degree of variability in PGSI consequence.

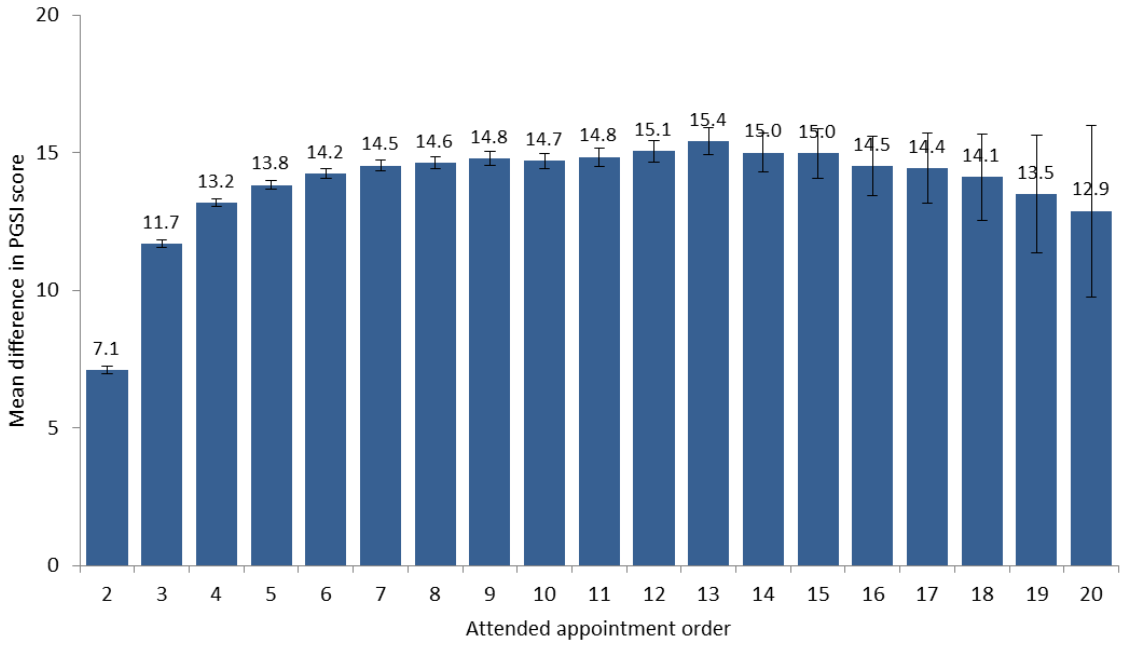
## 4.7 Gamblers' PGSI across all appointments

As with section 4.3 for the Core-10, this section uses the appointment level data to assess how treatment impacts PGSI scores at each attended appointment. The variables used here differ to previous sections as they are again based on the data from each individual appointment for each gambler.

We calculate difference in PGSI at each appointment by subtracting the initial baseline score (recorded at the first attended appointment) from the score at each attended appointment. This gives us a figure for PGSI difference at each appointment.

The greatest difference in PGSI score compared to the first appointment was at appointment 13, though as Figure 12 illustrates, there is only a small difference between the average at appointment 13 (M = 15.4, 95% CI [14.9,15.9]) and appointment 7 (M = 14.5, 95% CI [14.3,14.7]). The longer that treatment goes on, the higher the potential for service users to become treatment resistant. This would be one explanation for why after appointment 13, the mean difference in PGSI score starts to decrease, though the estimation uncertainty precludes drawing firm conclusions as to this trend. After appointment 6, there is overlap between the 95% CIs of almost all appointments.

*Figure 12: Mean difference in PGSI score (with 95% CI bars) by attended appointment order*




We tested the apparent non-linear impact of treatment using a mixed model, with the difference in PGSI compared to the first appointment PGSI as the dependent variable, and including both the attended appointment order and the squared attended appointment order as factors. The parameter estimates for difference in PGSI score by appointment are provided in Table 36.

The positive coefficient for the linear term and the negative coefficient for the quadratic term demonstrated that, like the Core-10 score, PGSI reduces a lot in the initial attended appointments before stabilising with further attended appointments. The maximum difference in PGSI score is attained at (keeping all other covariates fixed) $\frac{2.371}{2\times.108} = 11.0$ attended appointments.

*Table 36: Parameter Estimates of linear mixed model with quadratic effect on difference in PGSI Score*

| Parameter | Estimate | Std. Error | t | p value | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | 6.840 | 0.305 | 22.413 | <.001 | 6.242 | 7.439 |
| Attended appointment order | 2.371 | 0.024 | 100.607 | <.001 | 2.325 | 2.417 |
| Squared attended appointment order | -0.108 | 0.002 | -61.818 | <.001 | -0.111 | -0.105 |
| Previous session attendance | -0.127 | 0.055 | -2.294 | .022 | -0.236 | -0.019 |
| Completed Treatment | 1.049 | 0.143 | 7.334 | <.001 | 0.769 | 1.329 |
| Gender (Female) | -0.380 | 0.182 | -2.092 | .036 | -0.736 | -0.024 |
| Client age at referral | -0.039 | 0.006 | -6.277 | <.001 | -0.051 | -0.027 |
| Known treatment recurrence | -0.325 | 0.129 | -2.513 | .012 | -0.579 | -0.072 |
| Client with multiple episodes | -0.481 | 0.245 | -1.966 | .049 | -0.961 | -0.001 |
| Referral number (by client) | -1.579 | 0.171 | -9.253 | <.001 | -1.914 | -1.244 |

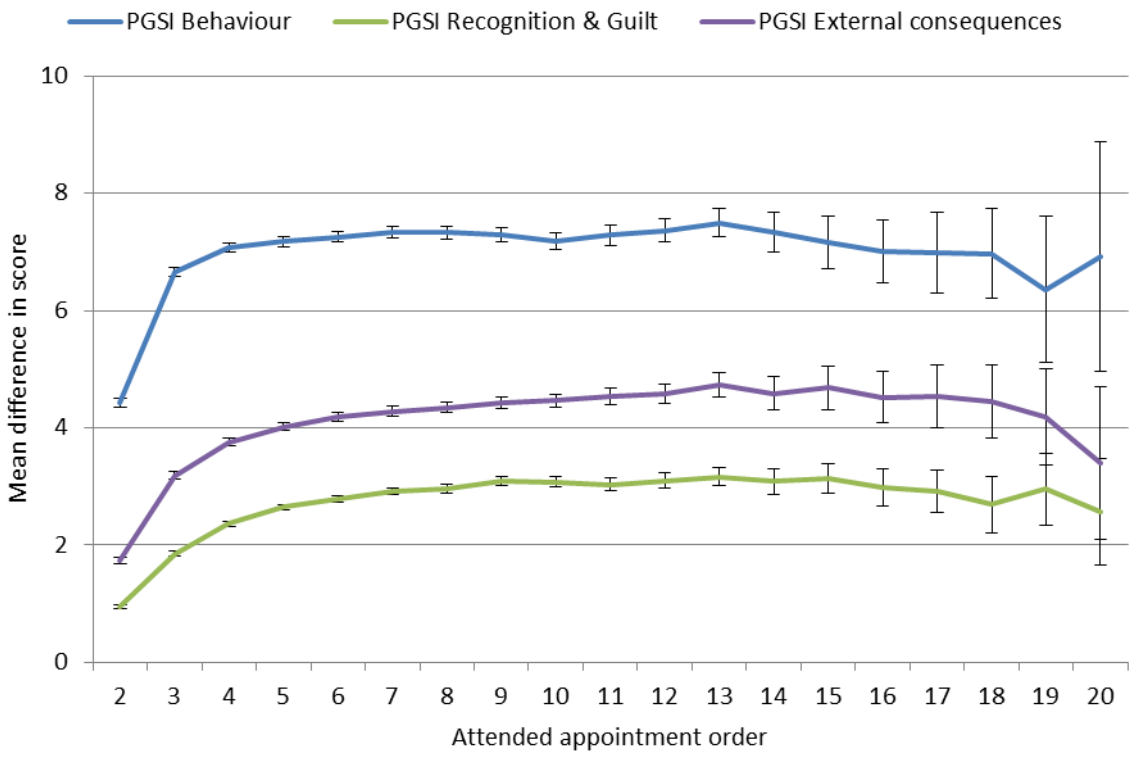
*Dependent variable (difference in PGSI score), N = 69,183*

For the linear mixed model there are no unique measure of effect sizes, so t statistic values were used to compare the relative effect sizes for the fixed factors. From the t values, the attended appointment order has the largest effect on the difference in PGSI score, followed by the quadratic attended appointment order. Difference in PGSI was also increased if the previous appointment was attended, illustrating the impact of missed appointments.

Where a gambler had repeat referrals, each additional referral (referral number) equated to a reduction in the difference in PGSI score of around 2 units (multiple referrals plus referral number). Completing treatment equated to an additional difference in PGSI score of 1.

As illustrated by Figure 13, the domains for PGSI appear to be impacted by treatment at different rates, with drastic change in the behaviour score early on in treatment, while both consequence domains have a much slower rate of change and remain at low levels of change throughout treatment. All domains show a decrease in difference beyond around 15 attended appointments, meaning the scores are closer to their initial baseline than previously, though again there is considerable estimation imprecision which precludes drawing firm conclusions on these data.

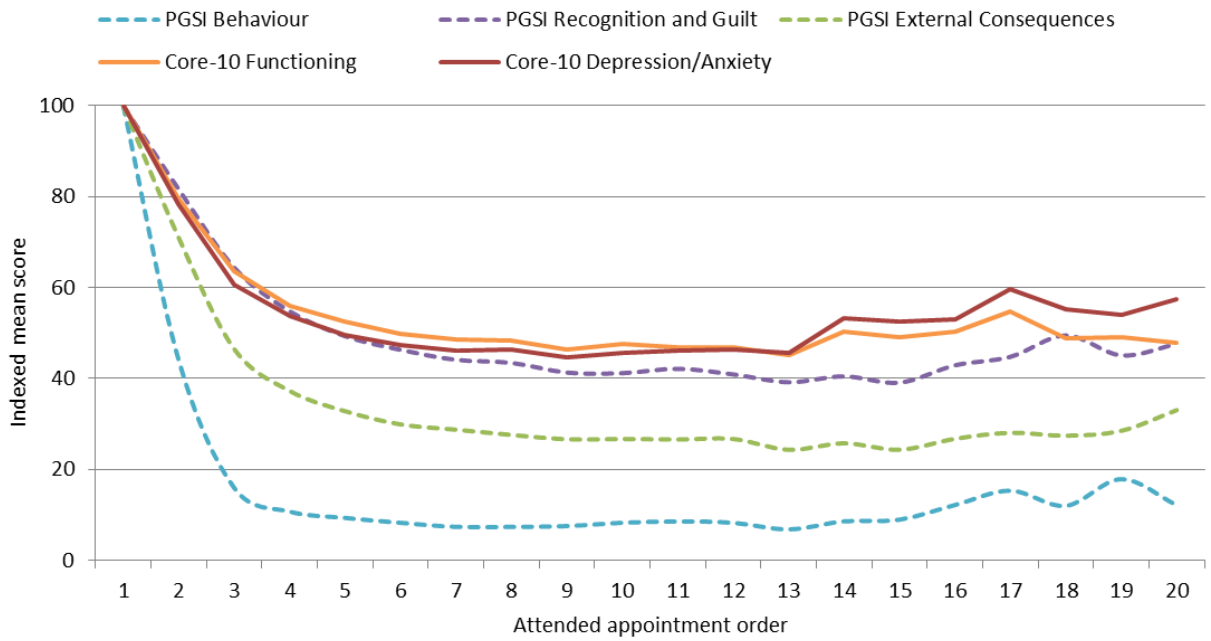
***Figure 13: Mean difference in PGSI domain scores compared to baseline (with 95% CI bars) by attended appointment order***




While there was some variation in how the domains of the gambler's Core-10 responded to treatment over time, there was less variation than observed in the PGSI domains. This is clearly illustrated when both the Core-10 and PGSI domain mean scores at each attended appointment are

indexed to the mean at the first appointment (Figure 14), which is equivalent to modelling the change compared to first appointment but removes the differences between domains based on maximum score. PGSI Behaviour shows a rapid decline and a consistently low score thereafter, with small increases later in treatment. PGSI recognition and guilt, however, more closely mirrors the Core-10 domains with a slower decrease, plateau and then a gradual but variable increase at later appointments. PGSI External Consequences decreases faster and farther than the Core-10 domains before stabilising but at a consistently higher level than PGSI behaviour.

***Figure 14: Indexed mean PGSI and Core-10 domain scores by attended appointment order for gamblers***




This is confirmed by the mixed modelling which shows the maximum difference in PGSI Behaviour score is attained at $\frac{.877}{2\times.045} = 9.8$ appointments, with parameter estimates provided in Table 37.

*Table 37: Parameter Estimates of linear mixed model with quadratic effect on difference in PGSI Behaviour Score*

| Parameter | Estimate | Std. Error | t | p value | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | 5.155 | 0.165 | 31.198 | <.001 | 4.831 | 5.479 |
| Attended appointment order | 0.877 | 0.012 | 73.969 | <.001 | 0.854 | 0.901 |
| Squared attended appointment order | -0.045 | 0.001 | -52.317 | <.001 | -0.046 | -0.043 |
| Previous appointment attendance | 0.053 | 0.029 | 1.852 | .064 | -0.003 | 0.109 |
| Completed treatment | 0.539 | 0.077 | 6.974 | <.001 | 0.388 | 0.691 |
| Gender (Female) | -0.124 | 0.099 | -1.252 | .211 | -0.317 | 0.070 |
| Service user age at referral | -0.017 | 0.003 | -5.108 | <.001 | -0.024 | -0.011 |
| Known treatment recurrence | -0.330 | 0.070 | -4.732 | <.001 | -0.467 | -0.193 |
| Service user with multiple referrals | 0.163 | 0.133 | 1.222 | .222 | -0.098 | 0.425 |
| Referral number (by service user) | -1.217 | 0.093 | -13.138 | <.001 | -1.399 | -1.036 |

*Dependent variable (difference in PGSI Behaviour score), N = 69,183*

The maximum difference in score for guilt and recognition is at $\frac{0.630}{2\times0.024} = 13.2$ appointments (Table 38), while for external consequences it is at $\frac{0.816}{2\times0.035} = 11.5$ attended appointments (Table 39).

*Table 38: Parameter Estimates of linear mixed model with quadratic effect on difference in PGSI Recognition and Guilt Score*

| Parameter | Estimate | Std. Error | t | p value | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | 0.407 | 0.082 | 4.980 | <.001 | 0.247 | 0.567 |
| Attended appointment order | 0.630 | 0.007 | 84.174 | <.001 | 0.615 | 0.645 |
| Squared attended appointment order | -0.024 | 0.001 | -43.168 | <.001 | -0.025 | -0.023 |
| Previous session attendance | -0.104 | 0.017 | -6.075 | <.001 | -0.138 | -0.070 |
| Completed Treatment | 0.264 | 0.039 | 6.754 | <.001 | 0.187 | 0.341 |
| Gender (Female) | -0.030 | 0.048 | -0.623 | .533 | -0.125 | 0.064 |
| Client age at referral | -0.007 | 0.002 | -4.455 | <.001 | -0.011 | -0.004 |
| Known treatment recurrence | -0.043 | 0.036 | -1.196 | .232 | -0.113 | 0.027 |
| Client with multiple episodes | -0.259 | 0.064 | -4.037 | <.001 | -0.385 | -0.133 |
| Referral number (by client) | -0.155 | 0.045 | -3.474 | <.001 | -0.242 | -0.068 |

*Dependent variable (difference in PGSI Recognition and Guilt score), N = 69,183*

Gender and known treatment recurrence are much weaker predictors for difference in PGSI Recognition and Guilt. Effects of all other variables on the scores are similar to the change in PGSI score model.

*Table 39: Parameter Estimates of linear mixed model with quadratic effect on difference in PGSI External Consequence Score*

| Parameter | Estimate | Std. Error | t | p value | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | 1.379 | 0.124 | 11.088 | <.001 | 1.135 | 1.622 |
| Attended appointment order | 0.816 | 0.010 | 84.656 | <.001 | 0.797 | 0.835 |
| Squared attended appointment order | -0.035 | 0.001 | -50.126 | <.001 | -0.037 | -0.034 |
| Previous session attendance | -0.083 | 0.023 | -3.629 | <.001 | -0.128 | -0.038 |
| Completed Treatment | 0.235 | 0.059 | 4.015 | <.001 | 0.121 | 0.350 |
| Gender (Female) | -0.206 | 0.074 | -2.769 | .006 | -0.352 | -0.060 |
| Client age at referral | -0.015 | 0.003 | -5.687 | <.001 | -0.020 | -0.010 |
| Known treatment recurrence | 0.039 | 0.053 | 0.739 | .46 | -0.065 | 0.143 |
| Client with multiple episodes | -0.362 | 0.100 | -3.610 | <.001 | -0.558 | -0.165 |
| Referral number (by client) | -0.200 | 0.069 | -2.912 | .004 | -0.334 | -0.065 |

*Dependent variable (difference in PGSI External Consequence score), N = 69,183*

Each of the models for the PGSI domains had a positive coefficient for the linear term and a negative coefficient for the quadratic term, indicating that although the domains are impacted by treatment differently over time, they all improve initially (faster in the case of Behaviour) and then stabilise with further attended appointments.

## 4.8 Reliable and clinically significant change in PGSI

In section 2.8, reliable change and clinically significant change for PGSI was defined. At treatment start, 94.5% of gamblers (11,664) met the threshold of a PGSI score of equal to or greater than 8 (referred to here as PG8) and 88.4% of gamblers (10,914) had a behaviour score of equal to or greater than 4 (referred to here as BH4).

*Table 40: Rates of reliable and clinically significant change in PGSI for gamblers using PG8 and BH4*

| | Using PG8 | | Using BH4 | |
|---|---|---|---|---|
| | N | % | N | % |
| Reliable and clinically significant improvement | 8,651 | 74.2% | 9,478 | 86.8% |
| Reliable improvement only | 1,887 | 16.2% | 478 | 4.4% |
| No reliable change | 1,014 | 8.7% | 868 | 8.0% |
| Reliable deterioration | 112 | 1.0% | 90 | 0.8% |
| Total | 11,664 | 100% | 10,914 | 100% |

Using PG8, 74.2% of all gamblers showed clinically significant improvement at treatment end. This was lower than when using BH4[32], where 87% of all gamblers showed clinically significant improvement at treatment end (Table 40).

The key difference between the two methods of assessing clinically significant change in PGSI, (PG8 and BH4) is that a higher proportion of gamblers who show reliable improvement also show clinically significant improvement using BH4. This is likely due to the lower reduction rates displayed within PGSI consequence scores (which would be included in PG8 but not in BH4), leading to a higher proportion of gamblers remaining above the 8-point threshold at treatment end.

As has been discussed at the end of section 4.6, there are legitimate reasons why the consequence elements of the PGSI may remain high and so it may be beneficial to assess the impact on just behaviour items when considering clinically significant change. However, there is a practical problem with using the behaviour item score to define clinically significant improvement, as it is not currently collected within the DRF framework and would place additional burden on data providers. This raises the question of whether it would be possible to improve the capture of gamblers who displayed reliable and clinically significant improvement in behaviour by adjusting the threshold for overall PGSI score.

As we increase the threshold score for PGSI overall, the proportion of gamblers with clinically significant improvement in behaviour who are also classified as being clinically significant via PGSI total score will increase. Alongside this there will also be a decrease in the number of gamblers without clinically significant behaviour change who also do not have clinically significant PGSI change. This could be thought of as sensitivity (true positive) and specificity (true negative)[33]; as the sensitivity increases, the specificity decreases and vice versa. A perfect model would have a sensitivity + specificity value of 2, while a useless model would have value of 1.

As shown in Table 41, increasing the PGSI score threshold increases the sensitivity but also quickly reduces the specificity, meaning more gamblers would be identified as having PGSI clinical change without having behaviour change. However, a single point adjustment to the threshold (so raising from >=8 to >=9) produces a marginally better sensitivity + specificity value, while maintaining very high specificity. The problem with using the screening threshold of >=8 is that it is not sensitive enough to those gamblers who show change in behaviour but maintain high guilt and problem

[32] It is important to note that only the qualifying criteria for clinical significance has changed, so the reliable change is still based on total PGSI score.

[33] Sensitivity relates to the question: what proportion of people who have clinically significant behaviour change also have clinically significant PGSI change? Specificity relates to the question: what proportion of people who do not have clinically significant behaviour change also do not have clinically significant PGSI change?

recognition scores. Increasing the threshold slightly would appear to go some way towards mitigating this issue.

***Table 41: Sensitivity and specificity values for BH4 when threshold for PG is adjusted***

| PG Threshold | Clinical BH4 = Clinical PG (sensitivity) | Not Clinical BH4 = Not Clinical PG (specificity) | Sensitivity + Specificity |
|---|---|---|---|
| >=6 | 76.6% | 99.8% | 1.764 |
| >=7 | 81.6% | 99.8% | 1.814 |
| >=8 (current) | 85.1% | 98.1% | 1.833 |
| >=9 | 88.2% | 96.2% | 1.845 |
| >=10 | 91.3% | 92.3% | 1.835 |
| >=11 | 93.2% | 87.4% | 1.806 |
| >=12 | 94.9% | 79.7% | 1.746 |

The current PGSI threshold has inherent shortcomings when used for assessment of change in gamblers in treatment. A single point shift in the threshold for the NGTS sample would incorporate more gamblers who had clinically significant change in behaviour, and thus may overcome some of these shortcomings. It is also worth noting that of the 9,956 gamblers that showed reliable improvement in PGSI (so ignoring the clinical threshold altogether), 95.2% of them had clinically significant change in PGSI behaviour scores. This leaves three possible options for assessing the impact of treatment, which we tested using logistic regression:

- Use reliable change plus a clinical threshold using the current threshold of >=8.
- Use reliable change plus a clinical threshold using the threshold of >=9.
- Use reliable improvement only.

Table 42 shows the outputs from the logistic regression model with clinically significant PGSI change in gamblers based on the original >=8 threshold as the dependent variable. The strongest effect was for brief referrals; gamblers with brief referrals had lower odds (Odds Ratio (OR) = 0.3, 95% CI [0.26, 0.34]) of showing clinically significant change in PGSI compared to those with longer referrals. High baseline scores for Core-10 Functioning and Core-10 Depression/Anxiety had a small negative impact on the odds of clinically significant change in PGSI.

***Table 42: Logistic regression output for clinically significant PGSI change (PG8) in gamblers (above PG8 at start of treatment)***

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Completed Treatment Status (completed) | 1.76 | < .001 | 1.57 | 1.97 |
| Total missed appointments | 0.89 | < .001 | 0.87 | 0.91 |
| Total attended appointments | 1.08 | < .001 | 1.06 | 1.10 |
| Brief referral (<=3 attended appts) | 0.30 | < .001 | 0.26 | 0.34 |
| Treatment recurrence | 0.73 | < .001 | 0.65 | 0.82 |
| Referral number | 0.73 | < .001 | 0.65 | 0.83 |
| Service user age at referral | 0.99 | .015 | 0.99 | 1.00 |
| Gender (Female) | 0.86 | .015 | 0.76 | 0.97 |
| Baseline Core-10 (Functioning) | 0.97 | .001 | 0.95 | 0.99 |
| Baseline Core-10 (Depression/Anxiety) | 0.94 | < .001 | 0.93 | 0.95 |
| Baseline Core-10 (Risk to Self) | 1.02 | .582 | 0.95 | 1.10 |
| Constant | 11.06 | < .001 | | |

*Dependent variable (clinically significant PGSI change PG8), N (gamblers above PG8 threshold at treatment start) = 11,664*

Table 43 shows the outputs from logistic regression with clinically significant PGSI change in gamblers based on >=9 threshold as the dependent variable. This was conducted on all gamblers who were >=9 PGSI score at treatment start. The impact of all variables was similar to the PG8 model and the confidence interval for all variables were very similar (and in the case of the Core-10 variables marginally smaller) than for the PG8 model.

***Table 43: Logistic regression output for clinically significant PGSI change (PG9) in gamblers (above PG9 at start of treatment)***

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Completed Treatment Status (completed) | 1.70 | < .001 | 1.51 | 1.92 |
| Total missed appointments | 0.87 | < .001 | 0.85 | 0.90 |
| Total attended appointments | 1.09 | < .001 | 1.06 | 1.11 |
| Brief referral (<=3 attended appts) | 0.28 | < .001 | 0.24 | 0.32 |
| Treatment recurrence | 0.70 | < .001 | 0.62 | 0.79 |
| Referral number | 0.74 | < .001 | 0.64 | 0.84 |
| Service user age at referral | 0.99 | .003 | 0.99 | 1.00 |
| Gender (Female) | 0.84 | .012 | 0.74 | 0.96 |
| Baseline Core-10 (Functioning) | 0.97 | < .001 | 0.95 | 0.98 |
| Baseline Core-10 (Depression/Anxiety) | 0.94 | < .001 | 0.93 | 0.95 |
| Baseline Core-10 (Risk to Self) | 0.99 | .804 | 0.91 | 1.07 |
| Constant | 17.37 | < .001 | | |

*Dependent variable (clinically significant PGSI change PG9), N (gamblers above PG9 threshold at treatment start) = 11,487*

Table 44 shows the outputs of the same model as the two above but using only reliable improvement in PGSI as the dependent variable. This was conducted on all gamblers who were >=9 PGSI score at treatment start, to ensure comparability with the above model.

***Table 44: Logistic regression output for reliable PGSI improvement in gamblers (above PG9 at start of treatment)***

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Completed Treatment Status (completed) | 1.57 | < .001 | 1.35 | 1.82 |
| Total missed appointments | 0.88 | < .001 | 0.85 | 0.91 |
| Total attended appointments | 1.11 | < .001 | 1.07 | 1.15 |
| Brief referral (<=3 attended appts) | 0.31 | < .001 | 0.26 | 0.38 |
| Treatment recurrence | 0.84 | .038 | 0.72 | 0.99 |
| Referral number | 0.66 | < .001 | 0.56 | 0.78 |
| Service user age at referral | 0.99 | < .001 | 0.98 | 0.99 |
| Gender (Female) | 0.84 | .051 | 0.71 | 1.00 |
| Baseline Core-10 (Functioning) | 0.98 | .062 | 0.95 | 1.00 |
| Baseline Core-10 (Depression/Anxiety) | 0.99 | .263 | 0.98 | 1.01 |
| Baseline Core-10 (Risk to Self) | 1.01 | .883 | 0.90 | 1.13 |
| Constant | 28.78 | < .001 | | |

*Dependent variable (reliable change in PGSI), N (gamblers above PG9 threshold at treatment start) = 11,487*

The models are again very similar, as is to be expected as clinically significant change is a subset of reliable improvement. The biggest difference is that the confidence intervals for all variables in the reliable improvement only model are much larger. This is because reliable improvement is a less stringent condition than clinically significant change and so potential error is likely to be greater.

The PGSI is not a validated outcome measure, but as it is the only gambling related measure consistently collected within the NGTS it is often used as a proxy outcome measure regardless. On a pragmatic level, it would be beneficial to improve the reporting of PGSI. Using a clinical threshold reduces the potential error in the data and helps to improve accuracy, without impacting the observed associations with treatment.

As shown earlier, raising the clinical threshold to >=9 more closely aligns with clinically significant change in PGSI behaviour at treatment end. It essentially mitigates some of the impact of PGSI recognition and guilt, which do not change as much as other measures during treatment. It also does not appear to impact the link between PGSI change and treatment in any meaningful way as shown by the logistic regression models. If it is at all appropriate to use a clinical threshold with PGSI data, then a score of >=9 would appear to be a better threshold to use within the NGTS.

The question, then, is does the additional requirement of a PGSI threshold for clinically significant change provide additional information that is helpful. One way to assess this would be to compare change in the PGSI with changes in Core-10. Are gamblers who show clinically significant change in PGSI (using >=9 threshold) more likely to show clinically significant change in Core-10 compared to

gamblers showing only reliable improvement in PGSI? This will be examined in the following section.

## 4.9 Comparing PGSI improvement with Core-10

It is commonly assumed that stopping problematic gambling (as indicated by reduction in PGSI score to below the problem gambling threshold) would be linked to improvements in Core-10. As discussed in the previous section there are two ways that we could assess overall improvements in PGSI: through reliable improvement only or reliable improvement with an end score below the threshold of a score of 9.

Table 45 shows the distribution of change in Core-10 by change in PGSI for gamblers who met both Core-10 threshold and PGSI (>=9) threshold at treatment start.

70.1% of gamblers who had clinically significant change in PGSI also had clinically significant change for Core-10. Conversely, 72.8% of gamblers who showed no reliable change or reliable deterioration in the PGSI also had no reliable change or reliable deterioration in Core-10.

***Table 45: Clinically significant change in Core-10 by clinically significant and reliable change in PGSI***

| | Clinically significant change in PGSI (PG9) | | Reliable improvement in PGSI (includes those with clinically significant) | | No reliable change/reliable deterioration in PGSI | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| Clinically significant improvement in Core-10 | 5,357 | 70.1% | 5,613 | 61.3% | 85 | 8.6% |
| Reliable improvement only in Core-10 | 1,100 | 14.4% | 1,632 | 17.8% | 183 | 18.5% |
| No reliable change/reliable deterioration in Core-10 | 1,189 | 15.6% | 1,914 | 20.9% | 719 | 72.8% |
| Total | 7,646 | | 9,159 | | 987 | |

*N (total gamblers who met both PGSI (PG9) and Core-10 thresholds) = 10,146*

61.3% of gamblers who had reliable improvement in PGSI had clinically significant improvement in Core-10. Again we see the impact of a less stringent condition increasing the variance within the group. Reliable improvement in PGSI contains more gamblers (including ones who did not make the PG9 threshold). The Core-10 scores for these gamblers are more variable than those gamblers who made clinically significant change. Clinically significant change in PGSI may therefore be a better predictor of clinically significant change in Core-10 than reliable improvement in PGSI on its own within our NGTS sample.

To test this, we ran two logistic regression models with Core-10 clinically significant change as the dependent variable. We included the same factors as in Table 31 but with the baseline PGSI scores removed. One model had clinically significant change in PGSI (PG9) as a variable and the other had

reliable improvement in PGSI. Both were applied only to gamblers who met both the Core-10 and PGSI clinical thresholds at the start of treatment to ensure consistency.

***Table 46: Logistic regression output for reliable and clinically significant Core-10 change in gamblers (clinical Core-10 and PG9 at treatment start)***

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Completed Treatment Status (completed) | 1.44 | < .001 | 1.27 | 1.63 |
| Total missed appointments | 0.89 | < .001 | 0.87 | 0.92 |
| Total attended appointments | 1.04 | < .001 | 1.02 | 1.05 |
| Brief referral (<=3 attended appts) | 0.51 | < .001 | 0.45 | 0.59 |
| Treatment recurrence | 0.75 | < .001 | 0.67 | 0.84 |
| Referral number | 0.94 | .366 | 0.82 | 1.08 |
| Service user age at referral | 1.00 | .019 | 0.99 | 1.00 |
| Gender (Female) | 0.67 | < .001 | 0.59 | 0.75 |
| PGSI Clinically Significant Change (PG9) | 11.09 | < .001 | 9.64 | 12.76 |
| Constant | 0.23 | < .001 | 0.00 | 0.00 |

*Dependent variable (clinically significant Core-10 change), N (total gamblers who met both PGSI (PG9) and Core-10 thresholds) = 10,146*

Gamblers who had made clinically significant improvement in PGSI (using PG9) had 11 times greater odds of also having made clinically significant improvement in Core-10 (OR = 11.09, 95% CI [9.64, 12.76]) than those who had not (Table 46).

***Table 47: Logistic regression output for reliable and clinically significant Core-10 change in gamblers (clinical Core-10 and PG9 at treatment start)***

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Completed Treatment Status (completed) | 1.57 | < .001 | 1.40 | 1.77 |
| Total missed appointments | 0.88 | < .001 | 0.86 | 0.90 |
| Total attended appointments | 1.04 | < .001 | 1.03 | 1.06 |
| Brief referral (<=3 attended appts) | 0.41 | < .001 | 0.36 | 0.47 |
| Treatment recurrence | 0.69 | < .001 | 0.62 | 0.77 |
| Referral number | 0.92 | .209 | 0.81 | 1.05 |
| Service user age at referral | 1.00 | .014 | 0.99 | 1.00 |
| Gender (Female) | 0.66 | < .001 | 0.59 | 0.75 |
| PGSI Reliable change | 10.80 | < .001 | 8.56 | 13.62 |
| Constant | 0.18 | < .001 | | |

*Dependent variable (clinically significant Core-10 change), N (total gamblers who met both PGSI (PG9) and Core-10 thresholds) = 10,146*

Gamblers who had made reliable improvement in PGSI also had around 11 times greater odds of having made clinically significant improvement in Core-10 (OR = 10.80, 95% CI [8.56, 13.62]) than those who had not (Table 47). However, there was also much greater uncertainty in the model, with larger confidence intervals than those of the clinically significant change model.

The PGSI model with an adjusted threshold for clinically significant change of =>9 at treatment start and <9 at treatment end provides a stronger and less uncertain predictor of Core-10 clinically significant improvement than simply reliable improvement in PGSI.

At treatment start, 93.0% of gamblers (11,486) met the threshold of a PGSI score of equal to or greater than 9 (PG9). Using PG9, 76.9% of all gamblers showed reliable and clinically significant improvement at treatment end (Table 48).

***Table 48: Rates of reliable and clinically significant change in PGSI for gamblers using PG9***

| | Using PG9 | |
|---|---|---|
| | **N** | **%** |
| Reliable and clinically significant improvement* | 8,827 | 76.9% |
| Reliable improvement only | 1,577 | 13.7% |
| No reliable change | 975 | 8.5% |
| Reliable deterioration | 107 | 0.9% |
| Total | 11,486 | 100.0% |

Typically, when evaluating the impact of something, we want to identify if it has worked well so that we can use these cases as case studies for best practice and see how they may differ from ones that are not as successful. Combining the new PGSI significant change with the Core-10 significant change data gives us a group of gamblers who showed significant change in both measures, and a group who showed no reliable change (or reliable deterioration) in both measures.

Of the 10,146 gamblers who met both thresholds at treatment start, 5,357 (52.8%) made clinically significant change in both PGSI and Core-10, while 736 gamblers (7.1%) made either no reliable change or reliable deterioration in both PGSI and Core-10.

We converted these two groups into two new variables and ran a logistic regression model for gamblers who met the thresholds for both PGSI and Core-10 at treatment start. Baseline scores for both PGSI and Core-10 were included in the model to test whether there were associations in the way that gamblers presented at treatment start and the outcome groups.

***Table 49: Logistic regression output for clinically significant Core-10 and PGSI change in gamblers (clinical Core-10 and PGSI at start excluding long referrals)***

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Completed Treatment Status (completed) | 1.78 | < .001 | 1.58 | 2.00 |
| Total missed appointments | 0.87 | < .001 | 0.85 | 0.89 |
| Total attended appointments | 1.05 | < .001 | 1.04 | 1.07 |
| Brief referral (<=3 attended appts) | 0.30 | < .001 | 0.26 | 0.34 |
| Treatment recurrence | 0.66 | < .001 | 0.60 | 0.74 |
| Referral number | 0.80 | < .001 | 0.70 | 0.90 |
| Service user age at referral | 0.99 | < .001 | 0.99 | 1.00 |
| Gender (Female) | 0.71 | < .001 | 0.63 | 0.80 |
| Baseline PGSI Score (Behavioural) | 1.00 | .806 | 0.98 | 1.02 |
| Baseline PGSI Score (Recognition and Guilt) | 1.10 | < .001 | 1.05 | 1.15 |
| Baseline PGSI Score (Other Consequences) | 0.96 | .004 | 0.94 | 0.99 |
| Baseline Core-10 (Functioning) | 0.99 | .445 | 0.98 | 1.01 |
| Baseline Core-10 (Depression/Anxiety) | 0.96 | < .001 | 0.95 | 0.97 |
| Baseline Core-10 (Risk to Self) | 0.99 | .689 | 0.92 | 1.06 |
| Constant | 2.40 | < .001 | | |

*Dependent variable (clinically significant Core-10 and PGSI change), N (total gamblers who met both PGSI (PG9) and Core-10 thresholds) = 10,146*

Table 49 shows the regression output for gamblers who made a clinically significant improvement in both PGSI and Core-10. Many of the same factors as were observed for PGSI and Core-10 improvement individually still applies, with completed treatment and attended appointments having a positive impact, and missed appointments and treatment recurrence having a negative impact. Female and older gamblers had lower odds of showing clinically significant imrpovement in both PGSI and Core-10, with female gamblers in particular having lower odds of clinically significant improvement in both compared to male gamblers by 29%.

Interestingly, higher PGSI recognition and guilt at the start of treatment equates to 10% higher odds of clinically significant improvement in PGSI and Core-10, while high PGSI external consequences reduced the odds by 4%. Higher Core-10 depression/anxiety scores at treatment start also reduced the odds by 4% that a gambler showed clinically significant improvement by treatment end.

One of the strongest associations was with brief referrals which reduced the odds of making clinically significant change in both PGSI and Core-10 to 30% compared to longer referrals.

*Table 50: Logistic regression output for no reliable change or reliable deterioration in Core-10 and PGSI change in gamblers (clinical Core-10 and PGSI at start excluding long referrals)*

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Completed Treatment Status (completed) | 0.64 | < .001 | 0.54 | 0.77 |
| Total missed appointments | 1.14 | < .001 | 1.09 | 1.19 |
| Total attended appointments | 0.90 | < .001 | 0.86 | 0.94 |
| Brief referral (<=3 attended appts) | 3.83 | < .001 | 2.99 | 4.92 |
| Treatment recurrence | 1.33 | .004 | 1.09 | 1.61 |
| Referral number | 1.29 | .018 | 1.05 | 1.59 |
| Service user age at referral | 1.01 | .159 | 1.00 | 1.01 |
| Gender (Female) | 1.28 | .021 | 1.04 | 1.57 |
| Baseline PGSI Score (Behavioural) | 0.91 | < .001 | 0.89 | 0.94 |
| Baseline PGSI Score (Recognition and Guilt) | 0.83 | < .001 | 0.77 | 0.90 |
| Baseline PGSI Score (Other Consequences) | 0.95 | .041 | 0.91 | 1.00 |
| Baseline Core-10 (Functioning) | 0.98 | .306 | 0.95 | 1.02 |
| Baseline Core-10 (Depression/Anxiety) | 1.01 | .42 | 0.99 | 1.03 |
| Baseline Core-10 (Risk to Self) | 0.99 | .929 | 0.87 | 1.14 |
| Constant | 0.34 | .001 | | |

*Dependent variable (no reliable change or reliable deterioration in Core-10 and PGSI), N (total gamblers who met both PGSI (PG9) and Core-10 thresholds) = 10,146*

Table 50 shows the regression output for gamblers who made no reliable improvement, or a deterioration, in both PGSI and Core-10. Age was not clearly associated with no reliable change, but female gamblers had almost 30% higher odds of showing no reliable improvement compared to male gamblers. Brief referrals had odds of no reliable change or reliable deterioration four times higher than longer referrals.

Higher starting scores for both PGSI behaviour and PGSI recognition and guilt equated to lower odds of no reliable change. It is likely there were two separate reasons for this. First, the effect of PGSI behaviour may be due to floor effects of low starting behaviour scores (e.g. those already abstaining can't show improvement in PGSI score due to already abstaining).

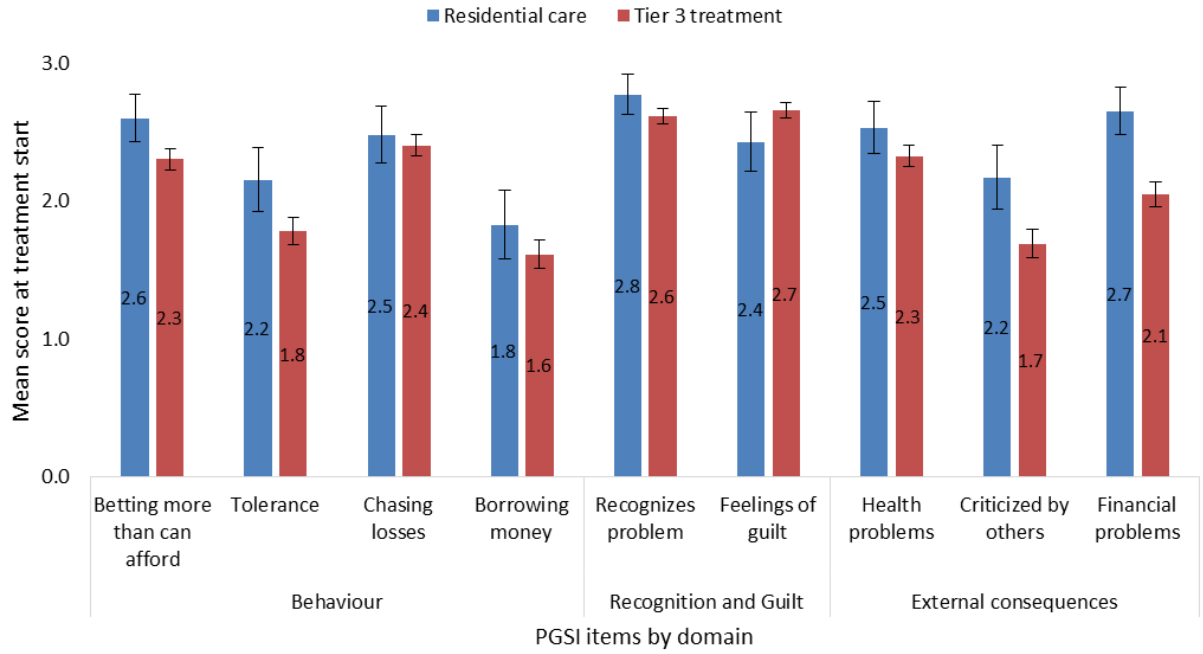
As PGSI recognition and guilt scores tended to change less during treatment, it is unlikely that this same explanation applied. Instead, the second explanation may be that higher starting scores for recognition and guilt indicate an initial acceptance of a problem which may make engagement in treatment higher, and therefore lead to lower likelihood of no reliable improvement. We will examine support for this hypothesis in section 5.4.

## 4.10 Change in gamblers PGSI for Tier 4 treatment

There were available data for 58 service users of residential care between 2018 and 2021, and 511 from Tier 3 treatment with comparable client type, gender and treatment length.

As shown in Figure 15, residential care service users had higher mean scores than Tier 3 service users at treatment start across all PGSI items except 'Feelings of guilt'. The overall PGSI score was higher amongst residential care service users with a mean of 21.6 (95% CI [20.7, 22.6]) compared to 19.5 (95% CI [19.0, 19.9]) for Tier 3 service users.[34] Baseline scores in PGSI Behaviour[35] and PGSI External Consequences[36] were also higher. There was no strong evidence for differences in PGSI Recognition and Guilt, with 'Feelings of guilt' the only item for which Tier 3 users (M=2.7, 95% CI [2.6, 2.7]) had higher mean starting score than residential care (M=2.4, 95% CI [2.2, 2.6]).[37]

***Figure 15: Mean PGSI item at treatment start (with 95% CI bars) by residential and Tier 3 treatment***




To compare Tier 3 treatment vs residential care, we used the difference in PGSI score between start and end of treatment (start minus end) as the dependent variable and treatment mode as the independent variable.

[34] T test t(567) = 3.120, p < .001

[35] T test t(567) = 2.335, p = .020

[36] T test t(567) = 4.328, p < .001

[37] T test t(567) = -2.421, p < .016

*Table 51: Regression output for the difference in PGSI score from start and end of treatment*

| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 16.31 | < .001 | 0.872 | 15.79 | 16.83 |
| Residential care | 3.50 | < .001 | 0.031 | 1.89 | 5.12 |
| Adjusted $R^2$ | 0.029 | | | | |

*Dependent variable (difference in PGSI score between start and end of treatment), N = 569*

Residential care had a small association with the difference in PGSI score compared to Tier 3 treatment (Table 51). Given the small effect size it is possible this is due to the higher starting scores. As residential care service users started with higher scores they have greater potential for improvement than Tier 3 service users.

*Table 52: Regression output for the difference in PGSI behaviour score from start and end of treatment*

| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 7.77 | < .001 | 0.848 | 7.49 | 8.04 |
| Residential care | 1.29 | .003 | 0.015 | 0.44 | 2.14 |
| Adjusted $R^2$ | 0.014 | | | | |

*Dependent variable (difference in PGSI Behaviour score between start and end of treatment), N = 569*

Similarly, residential care also had a small association with the difference in PGSI behaviour score (Table 52) and PGSI external consequences score (Table 53) compared to Tier 3 treatment. These were the same domains in which residential care service users had higher scores at treatment start.

*Table 53: Regression output for the difference in PGSI external consequences score from start and end of treatment*

| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 4.83 | < .001 | 0.742 | 4.593 | 5.06 |
| Residential care | 2.12 | < .001 | 0.054 | 1.385 | 2.855 |
| Adjusted $R^2$ | 0.052 | | | | |

*Dependent variable (difference in PGSI External Consequences score between start and end of treatment), N = 569*

In the only PGSI domain in which residential care service users did not have a higher mean starting score, Recognition and guilt, there was no clear difference compared with Tier 3 treatment (Table 54).

*Table 54: Regression output for the difference in PGSI recognition and guilt score from start and end of treatment*

| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 3.72 | < .001 | 0.760 | 3.54 | 3.89 |
| Residential care | 0.09 | .732 | <.001 | -0.45 | 0.63 |
| Adjusted $R^2$ | <.001 | | | | |

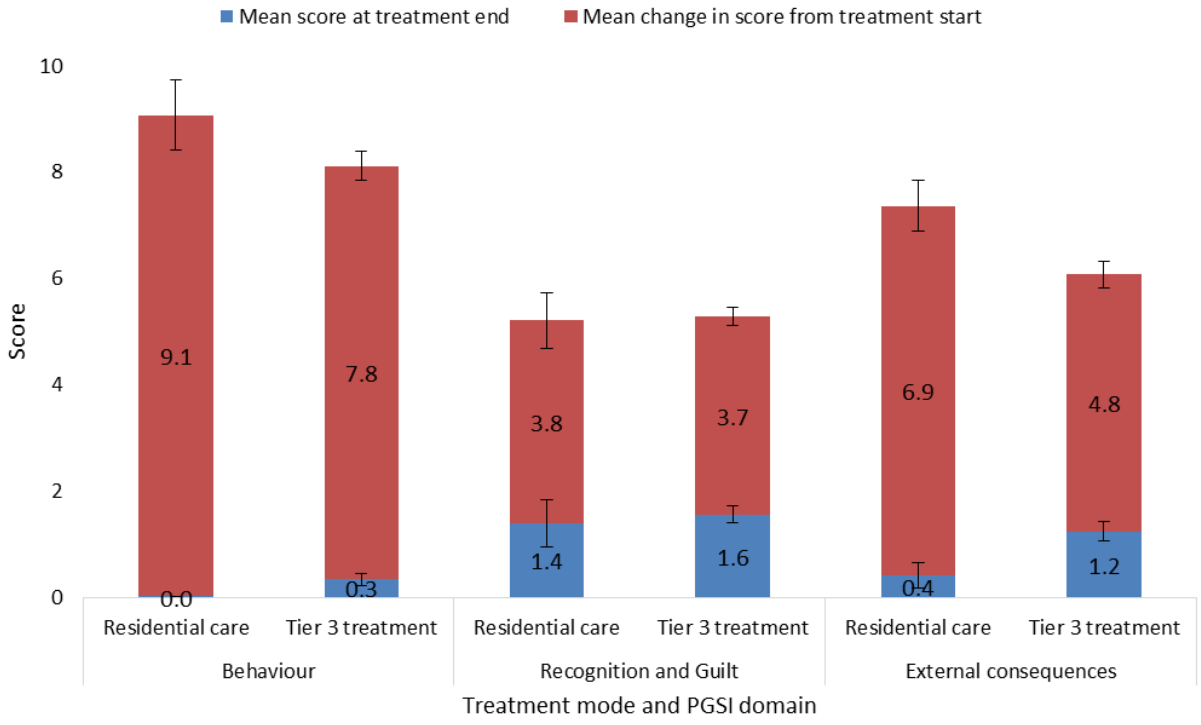
*Dependent variable (difference in PGSI Recognition and Guilt score between start and end of treatment), N = 569*

These findings are clearly illustrated in Figure 16 as behaviour showed a higher starting score for residential care but very low end scores for both treatment modes, recognition and guilt was very similar for both treatment modes, and external consequences showed residential with a higher starting score and lower end score than Tier 3 treatment.

At treatment end the mean PGSI score for residential care service users was 1.8 (95% CI [1.2, 2.4]) compared to 3.1 (95% CI [2.8, 3.5]) for Tier 3.

As gamblers in residential care would have no access to gambling they might also be expected to have lower behaviour scores at treatment end. This was evident in the data, with a mean end behaviour score for residential care service users of 0.02 (95% CI [0, 0.05]) compared to 0.34 (95% CI [0.23, 0.46]) for Tier 3 treatment. However, as the Tier 3 PGSI behaviour scores were also very low it is the improvement in PGSI external consequences (which focus on health problems, criticism from others and financial problems) which is arguably more interesting.

*Figure 16: Mean change and end treatment score (with 95% CI bars) by domain and treatment mode*



Residential care service users had a higher mean starting score for PGSI external consequences (M=7.4, 95% CI [7.0, 7.7]) compared to Tier 3 (M=6.1, 95% CI [5.9, 6.3]). However, at treatment end residential care users had lower mean PGSI external consequence scores (M=0.4, 95% CI [0.2, 0.7]) compared to Tier 3 (M=1.2, 95% CI [1.1, 1.4]), as shown in Figure 16.

This also has an impact on the number of service users ending treatment with reliable and clinically significant change in PGSI (Table 55). For residential care, 100% of service users had clinically significant change (using PG9), compared to 87.1% in Tier 3 treatment[38].

***Table 55: Rates of reliable and clinically significant change in PGSI for gamblers by Treatment mode***

| | Tier 3 | | Residential Care | |
|---|---|---|---|---|
| | **N** | **%** | **N** | **%** |
| Reliable and clinically significant improvement | 445 | 87.1% | 58 | 100.0% |
| Reliable improvement only | 45 | 8.8% | 0 | 0.0% |
| No reliable change | 21 | 4.1% | 0 | 0.0% |
| Reliable deterioration | 0 | 0.0% | 0 | 0.0% |
| Total | 511 | | 58 | |

Most of the difference between treatment modes was in whether the service user met the clinical threshold, with 95.9% of the Tier 3 service users showing reliable change in PGSI (irrespective of whether they met the clinical threshold or not). This may be due to the larger reduction in the PGSI external consequences, which make it more likely for a residential care service user to fall below the threshold of a score of 9.

[38] Chi squared test X2(1, 569) = 7.050, p = .008

# 5 Outcomes of Tier 3 treatment for affected others

## 5.1 Affected others' Core-10

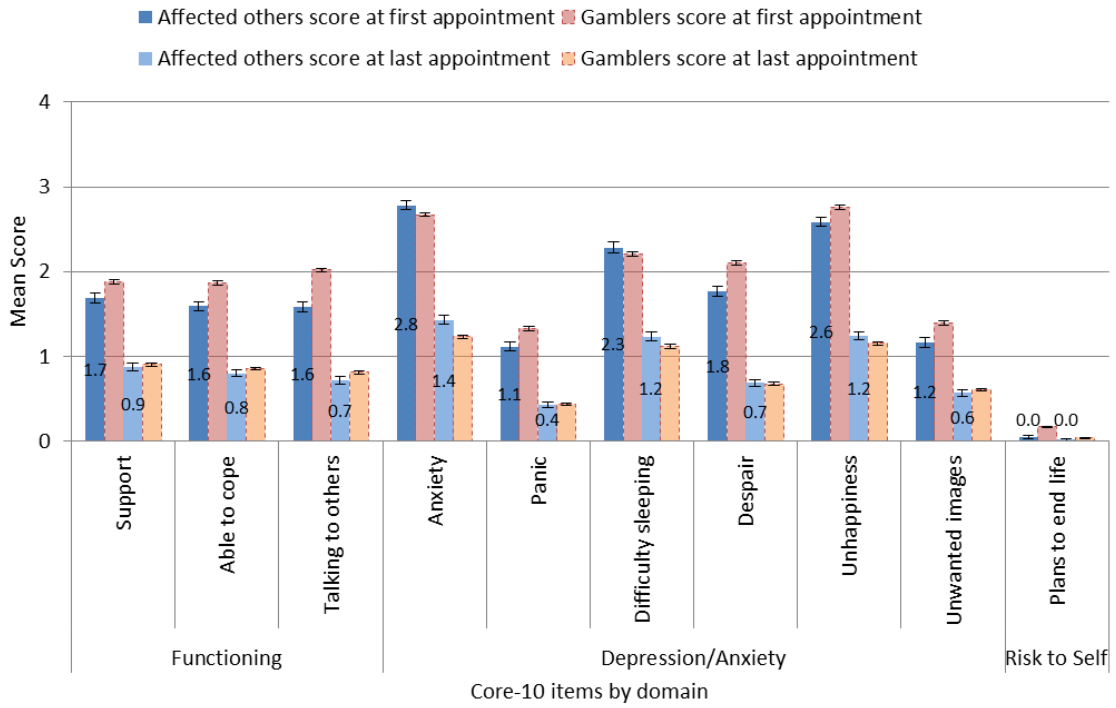
To assess the impact of treatment on affected others' Core-10 scores we measured the difference (first minus last) in the Core-10 score at the first attended appointment (typically an assessment) and at the last attended appointment at which a measure was available. Table 56 shows the pairwise t-test results to test the difference in mean scores. The results show that there were large reductions in the Core-10 total score and all domains between first and last attended appointment. The mean difference between total Core-10 score at first and last appointment is 8.6, which is lower than the reduction in gamblers (10.6). This may be due to a lower mean starting score, as affected others also had lower mean Core-10 score at first attended appointment (16.6) than gamblers (18.4).

***Table 56: Pairwise t-test results comparing Core-10 scores at first and last attended appointment for Affected others***

| | Paired difference between PGSI score at first and last appointment | | | |
|---|---|---|---|---|
| | | 95% CI of the Difference | | |
| | Mean difference | Lower | Upper | P-value |
| Core-10 Score | 8.61 | 8.28 | 8.94 | < .001 |
| Core-10 (Functioning) | 2.47 | 2.34 | 2.60 | < .001 |
| Core-10 (Depression and Anxiety) | 6.11 | 5.87 | 6.35 | < .001 |
| Core-10 (Risk to Self) | 0.03 | 0.02 | 0.05 | < .001 |

Figure 17 illustrates mean Core-10 item scores for affected others and gamblers at the first and last attended appointment. It is notable that affected others had higher scores in many of the depression and anxiety domain items than gamblers at treatment end. We will return to this in section 5.4.

*Figure 17: Mean Core-10 item score at last appointment and change from first for affected others and gamblers (with 95% CI bars)*




We performed multi-variable linear regression with the difference in Core-10 score in the first and last available appointment (first minus last), meaning a positive change value would show the affected others' Core-10 score reducing between first and last appointment, as detailed in section 2.10.

The model explains 10% of observed variance. Much like with the model for gamblers Core-10 (Table 24) the variable with the largest effect on difference in Core-10 was whether the affected other had a brief referral[39], with those who had showing smaller difference in PGSI.

Neither treatment completion nor, surprisingly, the number of attended appointments associated strongly with difference in Core-10, although missed appointments had the next largest effect after brief referrals. Female affected others had higher difference in Core-10 than male affected others, likely a result of the higher starting scores highlighted in Table 20.

[39] VIF = 1.829 so no issue of multi-collinearity with total attended appointments

*Table 57: Regression output for the difference in Core-10 scores at first and last appointment for affected others*

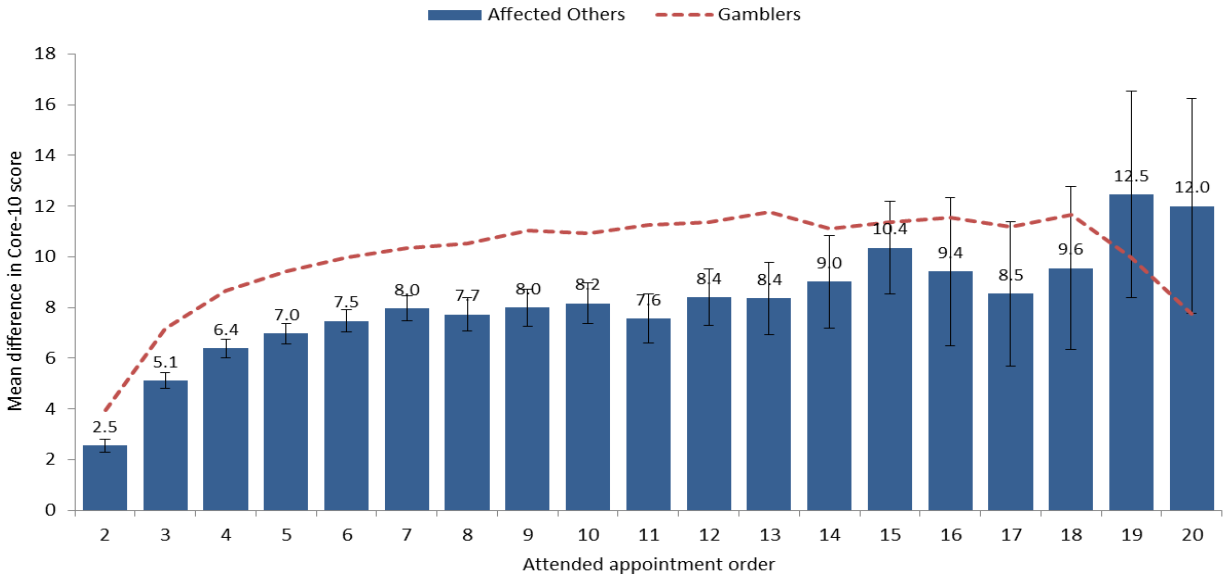
| Variable | Coefficient | p value | Effect Size | 95% CI | |
|---|---|---|---|---|---|
| | | | | Lower | Upper |
| Intercept | 8.92 | < .001 | 0.032 | 6.82 | 11.01 |
| Total attended appointments | 0.11 | .078 | 0.001 | -0.01 | 0.23 |
| Brief referral (<=3 attended appts) | -3.96 | < .001 | 0.030 | -4.92 | -3.00 |
| Total missed appointments | -0.55 | < .001 | 0.012 | -0.76 | -0.33 |
| Completed Treatment Status (completed) | 0.97 | .065 | 0.002 | -0.06 | 2.00 |
| Known treatment recurrence | 0.32 | .548 | 0.000 | -0.72 | 1.36 |
| Referral number | -1.22 | .005 | 0.004 | -2.06 | -0.37 |
| Gender (Female) | 2.01 | < .001 | 0.009 | 1.12 | 2.90 |
| Service user age at referral | -0.01 | .272 | 0.001 | -0.04 | 0.01 |
| Adjusted $R^2$ | 0.100 | | | | |

*Dependent variable (difference in Core-10 between first and last attended appointment), N = 2,114*

## 5.2 Affected others' Core-10 across all appointments

The greatest mean difference in Core-10 score compared to the first appointment for affected others was at appointment 19 (M = 12.5, 95% CI [8.4, 16.5]). Beyond appointment 13 though the mean scores were based on less than 100 affected others, so these results should be treated with caution as seen by the large confidence intervals (Figure 18). Before appointment 13, the largest difference in Core-10 score occurred at appointment 12, though there is little difference between the mean at appointment 7 (M = 8.0, 95% CI [7.5, 8.4]) and appointment 12 (M = 8.4, 95% CI [7.3, 9.5]). Affected others had lower mean difference in Core-10 than gamblers at almost every attended appointment.

*Figure 18: Mean difference in Core-10 score (with 95% CI bars) by attended appointment order for affected others*




We tested the apparent non-linear effect of treatment on Core-10 using a mixed model including both the attended appointment order and the squared attended appointment order as variables. The parameter estimates for difference in Core-10 score are provided in Table 58.

The positive coefficient for the linear term and the negative coefficient for the quadratic term indicate that the change in Core-10 increased with initial attended appointments before stabilising with further attended appointments.[40]

The lowest Core-10 score was attained at (keeping all other covariates fixed) $\frac{1.705}{2\times.057} = 14.9$ attended appointments which is rounded to 15. This was longer than for gamblers (14.1).

***Table 58: Parameter Estimates of linear mixed model with quadratic effect on difference in Core-10 score for affected others***

| Parameter | Estimate | Std. Error | t | p value | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | -2.020 | 0.782 | -2.583 | .01 | -3.555 | -0.486 |
| Attended appointment order | 1.705 | 0.059 | 28.966 | <.001 | 1.590 | 1.821 |
| Squared attended appointment order | -0.057 | 0.004 | -13.015 | <.001 | -0.066 | -0.049 |
| Previous session attendance | -0.392 | 0.139 | -2.826 | .005 | -0.665 | -0.120 |
| Completed Treatment | 0.761 | 0.389 | 1.956 | .051 | -0.002 | 1.524 |
| Gender (Female) | 1.222 | 0.381 | 3.210 | .001 | 0.475 | 1.968 |
| Client age at referral | -0.002 | 0.010 | -0.164 | .87 | -0.021 | 0.018 |
| Known treatment recurrence | -0.811 | 0.342 | -2.370 | .018 | -1.482 | -0.140 |
| Client with multiple episodes | -0.838 | 0.576 | -1.454 | .146 | -1.967 | 0.292 |
| Referral number (by client) | 0.961 | 0.404 | 2.375 | .018 | 0.166 | 1.756 |

*Dependent variable (difference in Core-10 score), N = 11,321*

The maximum difference in Core-10 Functioning score was attained at $\frac{0.479}{2\times0.015} = 15.8$ appointments (Table 59), while for Core-10 Depression/Anxiety it was attained at $\frac{1.217}{2\times0.044} = 14.0$ appointments (Table 60). Difference in Core-10 Functioning score is longer than for gamblers who had the greatest difference in Core-10 Functioning score at 14.3 appointments and Core-10 Depression/Anxiety at 13.9 appointments.

[40] As this model is identical to the one used for gamblers Core-10, it is notable that the affected others intercept is a negative number while the intercept for the gamblers model is positive. Fitting two identical models, one for gamblers and the other for affected others, is equivalent to fitting one single model with client type as a fixed factor and all other factors having interaction terms with the client type. The difference in the intercepts in the two separate models is exactly the same as the effect of the fixed factor of client type would be in a single model. This shows that change in Core-10 for gamblers is higher than change in Core-10 for affected others.

***Table 59: Parameter Estimates of linear mixed model with quadratic effect on difference in Core-10 functioning score for affected others***

| Parameter | Estimate | Std. Error | t | p value | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | -0.643 | 0.331 | -1.942 | .052 | -1.293 | 0.006 |
| Attended appointment order | 0.479 | 0.023 | 21.004 | <.001 | 0.435 | 0.524 |
| Squared attended appointment order | -0.015 | 0.002 | -8.962 | <.001 | -0.018 | -0.012 |
| Previous session attendance | -0.133 | 0.057 | -2.331 | .02 | -0.246 | -0.021 |
| Completed Treatment | 0.386 | 0.165 | 2.342 | .019 | 0.063 | 0.709 |
| Gender (Female) | 0.390 | 0.163 | 2.391 | .017 | 0.070 | 0.710 |
| Client age at referral | -0.002 | 0.004 | -0.482 | .63 | -0.010 | 0.006 |
| Known treatment recurrence | 0.138 | 0.143 | 0.962 | .336 | -0.143 | 0.418 |
| Client with multiple episodes | -0.328 | 0.249 | -1.315 | .189 | -0.816 | 0.161 |
| Referral number (by client) | 0.170 | 0.169 | 1.003 | .317 | -0.163 | 0.503 |

*Dependent variable (difference in Core-10 Functioning score), N = 11,321*

Indicators that the affected other had re-entered treatment (known treatment recurrence and referral number) had negative associations with the difference in Core-10 Depression/Anxiety, but were not strongly associated with changes in Core-10 Functioning. Associations of all other variables on the scores are similar to the change in Core-10 score model.

***Table 60: Parameter Estimates of linear mixed model with quadratic effect on difference in Core-10 depression/anxiety score for affected others***

| Parameter | Estimate | Std. Error | t | p value | 95% CI | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| Intercept | -1.511 | 0.598 | -2.527 | .012 | -2.684 | -0.338 |
| Attended appointment order | 1.217 | 0.044 | 27.735 | <.001 | 1.131 | 1.303 |
| Squared attended appointment order | -0.044 | 0.003 | -13.193 | <.001 | -0.050 | -0.037 |
| Previous session attendance | -0.277 | 0.104 | -2.668 | .008 | -0.481 | -0.074 |
| Completed Treatment | 0.421 | 0.296 | 1.423 | .155 | -0.159 | 1.002 |
| Gender (Female) | 0.843 | 0.290 | 2.904 | .004 | 0.274 | 1.413 |
| Client age at referral | 0.001 | 0.008 | 0.195 | .846 | -0.013 | 0.016 |
| Known treatment recurrence | -1.029 | 0.260 | -3.961 | <.001 | -1.538 | -0.520 |
| Client with multiple episodes | -0.597 | 0.441 | -1.354 | .176 | -1.463 | 0.268 |
| Referral number (by client) | 0.867 | 0.312 | 2.777 | .006 | 0.254 | 1.481 |

*Dependent variable (difference in Core-10 Depression/Anxiety score), N = 11,321*

The longer time required for the optimal impact on Core-10 functioning, which focuses on social support and talking to others, may be reflective of the isolation many affected others experience. Indeed, when compared with the mean Core-10 scores for gamblers, affected others had higher mean scores throughout most of the treatment despite lower scores in the first appointment. This will be explored in section 5.4.

## 5.3 Reliable and clinically significant change in affected others' Core-10

At treatment start, 81% of affected others (1,717) were above the clinical threshold of a Core-10 score of greater than or equal to 11, as shown earlier in Table 20. 52% of affected others who met the clinical threshold at treatment start showed reliable and clinically significant improvement in Core-10 at treatment end (Table 61).

***Table 61: Rates of reliable and clinically significant change in Core-10 for Affected Others***

| | Affected Others above Clinical threshold | |
|---|---|---|
| | **N** | **%** |
| Reliable and clinically significant improvement | 897 | 52.2% |
| Reliable improvement only | 313 | 18.2% |
| No reliable change | 477 | 27.8% |
| Reliable deterioration | 30 | 1.7% |
| Total | **1,717** | **100.0%** |

Table 62 shows the outputs from logistic regression with reliable and clinically significant Core-10 change as the dependent variable for affected others only. As with the model for gamblers, brief referrals had a strong association, reducing the odds of clinically significant change in Core-10 compared to full referrals by 66%.

***Table 62: Logistic regression output for reliable and clinically significant Core-10 change in affected others (clinical Core-10 at start)***

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Completed Treatment Status (completed) | 1.65 | .003 | 1.19 | 2.29 |
| Total missed appointments | 0.82 | < .001 | 0.77 | 0.88 |
| Total attended appointments | 1.00 | .93 | 0.96 | 1.04 |
| Brief referral (<=3 attended appts) | 0.34 | < .001 | 0.25 | 0.46 |
| Treatment recurrence | 0.77 | .098 | 0.56 | 1.05 |
| Referral number | 0.69 | .006 | 0.54 | 0.90 |
| Service user age at referral | 0.99 | .006 | 0.98 | 1.00 |
| Gender (Female) | 0.94 | .684 | 0.68 | 1.29 |
| Constant | 2.98 | .002 | | |

*Dependent variable (clinically significant Core-10 change), N (affected others clinical Core-10 at treatment start)= 1,717*

## 5.4 Comparing outcomes for gamblers and affected others

The analysis so far has indicated that while affected others may often start treatment with lower Core-10 scores than gamblers, they were quickly overtaken in terms of the speed of improvement in Core-10 and ended treatment with higher mean scores. This can be seen in the higher proportion of gamblers who were above the Core-10 threshold at treatment start (85% of gamblers compared to

81% of affected others) and the lower proportion of affected others who showed reliable and clinically significant change (52%) than gamblers (56%).

Affected others were two thirds (OR = 1.67, 95% CI [1.45, 1.93]) more likely to have a brief referral (<= 3 attended appointments) than gamblers, as shown in Table 63. The largest effect size was for treatment completion, with those who completed treatment five times less likely to have had a brief referral than those who dropped out.

Female service users and older service users both had lower odds of having brief referrals, as did those with high baseline Core-10 depression/anxiety scores at treatment start.

***Table 63: Logistic regression output for brief referrals***

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Client Type (Affected other) | 1.67 | < .001 | 1.45 | 1.93 |
| Completed Treatment Status (completed) | 0.20 | < .001 | 0.18 | 0.21 |
| Treatment recurrence | 0.93 | .154 | 0.83 | 1.03 |
| Referral number | 0.85 | .016 | 0.75 | 0.97 |
| Service user age at referral | 0.98 | < .001 | 0.98 | 0.99 |
| Gender (Female) | 0.82 | .001 | 0.73 | 0.92 |
| Baseline Core-10 (Functioning) | 0.99 | .086 | 0.97 | 1.00 |
| Baseline Core-10 (Depression/Anxiety) | 0.97 | < .001 | 0.96 | 0.98 |
| Baseline Core-10 (Risk to Self) | 0.97 | .496 | 0.90 | 1.05 |
| Constant | 4.04 | < .001 | | |

*Dependent variable (brief (<= 3 attended appointments) referrals), N (affected others and gamblers)= 14,462*

Affected others had 57% higher odds of completing treatment than gamblers, as shown in Table 64. Service users with brief referrals had 50% lower odds of completing treatment compared to those with longer referrals. High baseline Core-10 scores were associated with lower likelihood of treatment completion, Service user characteristics (gender, age, treatment recurrence) did not relate strongly to treatment completion.

*Table 64: Logistic regression output for treatment completion*

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Client Type (Affected other) | 1.57 | < .001 | 1.32 | 1.88 |
| Total missed appointments | 0.65 | < .001 | 0.63 | 0.66 |
| Total attended appointments | 1.30 | < .001 | 1.27 | 1.34 |
| Brief referral (<=3 attended appts) | 0.50 | < .001 | 0.44 | 0.57 |
| Treatment recurrence | 0.94 | .288 | 0.83 | 1.06 |
| Referral number | 1.30 | .002 | 1.10 | 1.54 |
| Service user age at referral | 1.00 | .319 | 1.00 | 1.01 |
| Gender (Female) | 0.97 | .614 | 0.85 | 1.10 |
| Baseline Core-10 (Functioning) | 0.96 | < .001 | 0.95 | 0.98 |
| Baseline Core-10 (Depression/Anxiety) | 0.99 | .021 | 0.98 | 1.00 |
| Baseline Core-10 (Risk to Self) | 0.88 | .003 | 0.81 | 0.96 |
| Constant | 2.91 | < .001 | | |

*Dependent variable (treatment completion), N (affected others and gamblers)= 14,462*

Interestingly, despite evidence showing that affected others had a lower change in Core-10 score than gamblers, there were no clear differences between client type for achieving clinically significant change in Core-10 by treatment end (Table 65). This may have been a result of many affected others being above the Core-10 threshold but with lower mean starting scores in Core-10 than gamblers. This would mean their Core-10 scores would not need to reduce as much to be clinically significant as gamblers. Brief referrals were less than likely to achieve clinically significant change in Core-10 compared to longer referrals by 69%.

*Table 65: Logistic regression output for clinically significant change in Core-10 (above Core-10 threshold)*

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Client Type (Affected other) | 1.01 | .908 | 0.88 | 1.15 |
| Total missed appointments | 0.84 | < .001 | 0.82 | 0.86 |
| Total attended appointments | 1.05 | < .001 | 1.04 | 1.07 |
| Brief referral (<=3 attended appts) | 0.31 | < .001 | 0.27 | 0.34 |
| Treatment recurrence | 0.70 | < .001 | 0.63 | 0.77 |
| Referral number | 0.83 | < .001 | 0.75 | 0.92 |
| Service user age at referral | 0.99 | < .001 | 0.99 | 1.00 |
| Gender (Female) | 0.70 | < .001 | 0.63 | 0.77 |
| Constant | 3.26 | < .001 | | |

*Dependent variable (clinically significant change in Core-10), N (affected others and gamblers above Core-10 clinical threshold at treatment start)= 12,230*

Brief referrals were almost three times more likely to show no reliable change or reliable deterioration in Core-10 than longer referrals (

Table 66). Female service users were also more likely to show no reliable change or reliable deterioration compared to male service users by 24%.

***Table 66: Logistic regression output for no reliable change or reliable deterioration in Core-10 (above Core-10 threshold)***

| Variable | OR | P-value | 95% CI | |
|---|---|---|---|---|
| | | | Lower | Upper |
| Client Type (Affected other) | 1.15 | .063 | 0.99 | 1.33 |
| Total missed appointments | 1.15 | < .001 | 1.13 | 1.18 |
| Total attended appointments | 0.94 | < .001 | 0.93 | 0.96 |
| Brief referral (<=3 attended appts) | 2.89 | < .001 | 2.56 | 3.27 |
| Treatment recurrence | 1.29 | < .001 | 1.16 | 1.43 |
| Referral number | 1.24 | < .001 | 1.12 | 1.39 |
| Service user age at referral | 1.00 | .662 | 1.00 | 1.01 |
| Gender (Female) | 1.24 | < .001 | 1.11 | 1.39 |
| Constant | 0.20 | < .001 | | |

*Dependent variable (no reliable change or reliable deterioration in Core-10), N (affected others and gamblers above Core-10 clinical threshold at treatment start)= 12,230.*

# 6 Conclusions and recommendations

## 6.1 Reporting treatment outcomes

Our analysis found that when gamblers engaged in treatment (i.e. attended more appointments), they had lower Core-10 and PGSI scores; 90.6% and 73.7% of gamblers showed reliable change in PGSI and Core-10 score between first and last appointment respectively. We also observed several factors that associated with the likelihood of gamblers making reliable and clinically significant change in both PGSI and Core-10. These included factors related to the referrals (attended appointments, treatment completion) and the gambler (gender, treatment recurrence and severity at start of treatment). Understanding what is driving these differences in outcomes requires consistent and appropriate outcome reporting.

Our results contradict the annual NGTS statistics publication (GambleAware, 2021) which states that 16.8% of service users show no change in PGSI or Core-10 score appointment. We found that only 2.0% of gamblers had the same PGSI score at first and last appointment, and only 3.3% of all service users had the same Core-10 score. This is likely because measures of change between earliest and latest appointment in the NGTS statistics publication include service users who had only one appointment. It may be beneficial to report data on service users who drop out pre-treatment separately from change in PGSI and Core-10 of those engaging in treatment.

Our results suggest potential benefits of using a PGSI threshold of <9 instead of <8 for clinically significant change and a reliable change of >=5 when measuring treatment effectiveness. These appear appropriate thresholds within the NGTS, though further research with other datasets is required to confirm the broad suitability of this approach. Reporting changes in the Core-10 is simpler than for PGSI as it is a validated outcome measure with agreed clinical cut-offs. Reporting service users who had made reliable and clinically significant change in Core-10 and those who had made no reliable change or showed reliable deterioration in Core-10 would be useful indicators of treatment impact.

Finally, there may be benefit in comparing gamblers who had made reliable and clinically significant change in both Core-10 and PGSI with gamblers who did not. Future research may examine how accurately these combined cut-offs identify high and low risk clients, and client treatment response.

It is important to note with reporting of PGSI at treatment end that the PGSI is not a clinical tool or outcome measure of treatment[41], even though it is commonly used as one (c.f. NGTS treatment statistics; GambleAware, 2021). While continued use of the PGSI would allow comparable data over

[41] GambleAware, 2021, footnote 15

time, a new validated gambling related measure would benefit assessment of treatment effectiveness. The Gambling Disorder Identification Test (GDIT)[42] was recently developed in an international Delphi and consensus process (Molander et al., 2021). This measure has the benefit of meeting the requirements of the Banff Consensus Agreement on gambling related measures (Walker et al, 2006) which the PGSI does not.

Many GDIT items are already collected in the DRF in some form; a complete comparison of the data items in GDIT and currently collected in DRF provided in Annex 4. Minor amendments to existing items and the introduction of four new questions, as described in section 6.8, would improve existing data collection in the DRF, minimise new question additions, and be a step towards the introduction of a validated gambling related outcome measure which may overcome the shortcomings of the PGSI.

## 6.2 Treatment length

On average, service users within the NGTS attended around 6 appointments during a referral. But this obscures wide variation in treatment journeys, with 24% of service users attending 3 or fewer appointments, and 21% attending 10 or more.

Existing literature suggests there is little evidence to connect positive outcomes with any specific number of sessions or length of treatment. Ribeiro et al. (2021), for example, concluded that, in relation to CBT, longer interventions seem to be no more beneficial than shorter ones. In similar vein, Parker and Bauermann's (2015) meta-analysis concluded that the effectiveness of psychological interventions was unrelated to the number of sessions.

Our research found that service users who attended 3 or fewer appointments had worse outcomes in terms of improvement in PGSI and Core-10 score and lower levels of clinically significant change for both measures. This does not mean that these brief referrals had no beneficial impact, but that their impact was less beneficial than longer treatments.

This may be due to the PGSI and Core-10 measures, and items within these, responding to treatment at different rates. Core-10 responded to treatment slower than the PGSI (with change maximised at 14 attended appointments for Core-10 compared to 12 for PGSI but variable after around 6 attended appointments for both measures) and we observed no improvements beyond around 14 appointments. The PGSI behaviour domain reduced quickly and remained low throughout treatment in the NGTS sample. PGSI consequence reduced more slowly, and this may be because it is constructed of items relating to recognition of guilt and external consequences

[42] Publicly available at https://gditscale.com/

which may take longer to address than changes in behaviour that can be implemented immediately (e.g. installing blocking software).

The difference in outcome of brief referrals may also reflect a lack of engagement on the part of the service user, which is known to have an impact on treatment efficacy via increased dropout (Dowling et al., 2021).

## 6.3 Treatment dropout and completion

Treatment completion is important as treatment effectiveness is impacted by low take-up of treatment and dropout (Gunstone and Gosschalk, 2020; Kalbfleisch et al., 2020). Our analysis found that treatment completion had a positive association with improvements in both PGSI and Core-10, with gamblers who completed treatment 78% more likely to have clinically significant change in both PGSI and Core-10 than those who dropped out. Service users with high Core-10 scores at treatment start were less likely to complete treatment, though there was no evidence that age or gender associated with completing treatment.

A lack of social support, particularly in situations of stress, has previously been found to associate with dropout from psychological treatment for substance abuse (Sayreet al., 2002). Other studies have found that service users in gambling treatment without other traditional social support like employment (Hodgins et al. 2004) or a relationship (Aragay et al. 2015) were more likely to drop out. Our analysis found that high Core-10 Functioning scores were predictive of higher rates of drop-out, which potentially ties into this existing research as they are indicative of struggling to maintain meaningful social support.

How drop out is defined varies in the literature with typically one of two approaches taken. The first is based on the therapist's judgement of appropriate termination, while the second is based on attendance at a specific number of appointments (Melville et al., 2007). The current approach taken in the NGTS is the former, as treatment within the NGTS is typically of an unspecified duration. Comparisons are further complicated by differences in the stage of drop-out, with some studies drawing distinctions between pre-treatment drop out and in-treatment drop-out (Ronzitti et al., 2017). Our analysis only included service users who engaged in more than one attended appointment, but there were a sizable number of NGTS service users who attended a single appointment (typically an assessment) and did not return to treatment thereafter. Comparisons would be improved by a more detailed approach to the definition of drop-out that distinguishes between drop-out at different stages of treatment.

## 6.4 Treatment recurrence

On therapeutic interventions, Blank et al. (2021) conclude that despite the ‘considerable number’ of therapeutic interventions for gambling, the evidence points only to short-term positive outcomes.

There is little evidence on the longer-term impact or to support one kind of therapy or mode of delivery over another. One of the gaps they identified was evidence of ongoing support after treatment given that "well over half of all incident problem-gambling cases are previous problem gamblers who are relapsing"[43].

Nearly one quarter (24.1%) of all NGTS service users in our sample had known treatment recurrence, and our analysis demonstrated that this was associated with how a service user responded to treatment. Yet there is little evidence on the availability of ongoing structured support within the NGTS. The DSM-5 specifies that early remission is a situation whereby an individual who previously met the full criteria for gambling disorder has now met none of the criteria for at least three months but for less than 12 months; sustained remission would apply when none of the criteria had been met for 12 months or more.

The DRF collects data on appointments with a purpose of "Follow-up after treatment", but of the 14,462 referrals we analysed, only 8.9% had any attended appointments with a purpose of "Follow-up after treatment". It is not clear if these are structured or informal follow-ups. In the latest DRF specification (GambleAware, 2021b) there are additional categories for "aftercare" and "Formal structured follow-up", with the wording on the latter amended to clarify that it should only apply to formal follow-up appointments. These amendments suggest a recognition that the NGTS would benefit from a better understanding of the post-treatment offers.

For service users with such a high chance of relapse, it is better that they seek help than fall back into harmful gambling behaviours. Treatment recurrence should not automatically be seen as a failure of treatment, even when it may reduce the impact of future treatment, and nor should maintaining a service user in treatment always be seen as a benefit (Walker, 2009). The willingness of a service user to reengage with treatment is in many ways a positive (North et al., 2014). The NGTS should have a clear and accessible policy towards service users returning to treatment, and how this is both complimented by, and distinguished from, aftercare.

## 6.5 Comparing Tier 3 and Tier 4 treatment

Gamblers in Tier 4 residential care treatment had higher starting PGSI scores compared to a group of comparable Tier 3 gamblers (mean of 21.6 compared to 19.5). Tier 4 residential care treatment associated with greater reductions in PGSI total score compared to Tier 3, with 100% of gamblers showing clinically significant reductions in PGSI compared to 87.1% in the Tier 3 group. This difference in outcome was driven mainly by greater reductions in PGSI external consequence scores (items focused on health problems, criticism from others and financial problems).

[43] Blank et al. (2021), e62

Due to a lack of reliable information about what interventions were used within a referral, caution should be exercised when drawing conclusions from comparisons between referrals. Tier 3 in the context of NGTS is wide ranging and potentially incorporates CBT, face-to-face and remote treatment, couples and group therapy, motivational interviewing and more. The DRF attempts to collect intervention used at appointment level, but this is almost exclusively categorised as general "counselling". It is therefore difficult to distinguish when multiple interventions have been used in an appointment. Comparisons are further complicated as the current structure of the DRF, particularly the appointments table, is not suited to capturing reliable data from Tier 4 residential care or 'retreat and counselling' services.

## 6.6 Affected others

There is ample existing evidence of the harms that another person's gambling can cause (Goodwin et al., 2017; Nash et al., 2018) but affected others made up only 14.6% of the NGTS service users in our sample. Our analysis found that although affected others had a lower change in Core-10 score than gamblers, they were no more likely than gamblers to achieve a clinically significant change in Core-10 by treatment end. Affected others were more likely to have brief referrals and to complete treatment than gamblers.

Kourgiantakis et al. (2013) found that treatment for affected others was linked to better coping skills and improved family functioning, even when the gambler was not also in treatment. Gamblers who entered treatment or couples therapy with a partner showed lower frequency of gambling urges and greater control in maintaining abstinence (Lee, 2002) and had higher attendance and lower likelihood of dropout (Jimenez-Murcia et al., 2017). Nilsson et al. (2020) found no differences in gambler outcomes but found that treatment take-up of online CBT was higher when it involved a partner.

In their analysis of affected others attending a problem gambling treatment clinic in London, Orford et al. (2017) found that around half (50.5%) of all affected others were spouses or partners of a gambler, with a further third (37.4%) a parent. The impact that gambling had on an affected other varied depending on the relationship to the gambler and whether they were living together or had regular contact.

The DRF does not collect information on affected others' relationship to the gambler, whether they live together, or whether the gambler is also in treatment. Collecting this data could therefore create opportunities for future research.

## 6.7 Limitations and future work

There was a lack of reliable and detailed data on what interventions were used with service users by each provider. This meant that potentially very different treatment approaches were grouped under a single “Tier 3” category. Development of more appropriate ways to capture intervention data would improve the value of the DRF data when comparing treatment impact.

As has already been discussed, the appointments data from Gordon Moody, which is the data container within the DRF for PGSI and Core-10 measures, is not suitable for Tier 4 residential care. This meant that the possible analysis of Tier 4 treatment was very limited, and excluded any data on Retreat and Counselling Services. Further work on the observed differences between Tier 4 and Tier 3, with particular focus on how Tier 4 treatment approaches managing external consequences may lead to beneficial shared learning opportunities. This work would also be complimented by additional data on the longevity of outcomes that comparable follow up data would provide (as discussed in section 6.4 above) particularly as residential care service users are in a somewhat artificial environment when they complete treatment.

The DRF data required a lot of cleaning prior to analysis, which would make future comparative work harder to accomplish without detailed and comprehensive guidance and coding documentation. This should be published alongside the data framework specification. The DRF data that is made available for external research should be assessed and potentially have additional processing to make future research using the data more efficient and ensure comparability.

Collecting measures at each appointment provides a rich source of data, but the measures do not change much between appointments once the service user is established in treatment. A review of the cost/benefit of collecting measures at each appointment would be advisable.

One of the required data cleaning decisions was removing people who had more than 20 appointments (who also had higher Core-10 scores). Inevitably this is likely to result in a sampling bias as we have removed some of the people with the worst mental health. As the numbers involved are very small this is unlikely to impact our findings but it is worth noting. It was outside the scope of this project to examine the reasons why some service users stay in treatment for extended periods (there are service users in the DRF data who continue having treatment over years). These small number of cases would benefit from a more qualitative approach with input directly from the practitioners themselves to identify why some service users stay in treatment for much longer than others. Anecdotally, many of these cases are likely to have safeguarding concerns or more complex requirements. Future research into how these groups are best supported within the NGTS network would be beneficial.

There were 462 referrals where the final attended appointment had no available measure. Instead of removing them from the analysis we assessed total attended measures up to the point that the

last available measure is available, as there is no way to assess the impact of any attended appointments after this point. As treatment generally stabilises in outcomes and there was no evidence of substantial differences between the last and second to last appointment outcomes we felt this was more proportionate than discarding the data all together. However, there is always a risk that this resulted in potentially biased results in these final appointments.

As discussed in Section 6.1, the introduction of a new gambling related measure would be beneficial to the NGTS. This should not be rushed and further testing would be helpful, both to identify the best measure going forward, and to ensure the greatest value from existing PGSI data. Changes to existing DRF data may enable testing of a proxy for GDIT without introducing the entire measure. This may also help to validate a clinical threshold for PGSI. Consideration should also be given to the reference period for any measure as it is typically 12 months but this may not make sense for service users in treatment.

This research examined the changes in measures across the entire treatment journey, with the focus on measures while the service user was in treatment. Understanding how improvements in outcomes during gambling treatment relate to longer-term sustainable remission of harmful behaviours is key to understanding the true impact of the NGTS. Future research should examine the existing follow-ups data within the DRF for both insight into impact and also ways to improve coverage and quality.

The data that was selected from the DRF was purposefully limited due to the scope and complexity of the appointments data we wanted to examine. In hindsight, additional variables on gambling behaviours (types of gambling, monthly expenditure) and social support structures (i.e. relationship status, employment status and dependent children) could have been included in the modelling without too much additional complexity and would have added valuable depth to the models relating to outcomes for gamblers.

The decision to analyse gamblers and affected others separately (with the aim of enabling more accurate modelling between Core-10 and PGSI) meant that some of the conclusions that could be drawn from the affected others analysis were more complicated than they needed to be (the results would have been the same but we could have approached the modelling in a simpler, more user friendly way).

Finally, as with any research that uses multiple outcome measures at different points in time, it is worth highlighting that much of this work hinges on an assumption of longitudinal measurement invariance: that the repeated measures (Core-10 and PGSI) represent the same thing to the service user over time. Part of the reason the measures are validated is to reduce the risk of changes in interpretation or meaning. But there is always a risk that these measures may mean different things to different people at different points during treatment.

## 6.8 Recommendations

Based on the analysis and findings of this project, the following recommendations are made regarding the collection and analysis of DRF data from NGTS providers:

### Outcome measures

- Separately report non-attendance of treatment and exclude service users who do not attend multiple sessions from calculations of outcome measure change. This includes service users who either attended a single assessment appointment or agreed to treatment but then failed to attend any further appointments.
- Use reliable and clinically significant change as the primary measure of change for the Core-10.
- Move away from using the screening threshold of <8 PGSI score when assessing the effectiveness of gambling treatment. Further work should be undertaken to establish a suitable threshold for clinically significant change in PGSI, with an initial position of reliable change in PGSI equating to change of >=5 and clinically significant change using a <9 threshold.
- Include a calculation of service users who make no reliable change or reliable deterioration in both Core-10 and PGSI when reporting on treatment impact for the NGTS.
- Consider the addition of four individual questions (GDIT items 3, 4, 7 and 9) and the amendment of four existing DRF items (relationship loss, job loss, time spent gambling, and daily time spent gambling) to correspond to the GDIT. This will enable comparison of the PGSI with elements of the GDIT, improve existing collections and having minimal impact on burden. The PGSI questions should remain the same as any changes to these would result in invalidating the comparability of the measure data and introduce longitudinal measurement variance (i.e. measuring something differently over time).
- Review whether PGSI and Core-10 collection at every appointment is necessary. This may result in a loss of data on service users who drop out so wider consultation would be needed on the potential benefits and risks.

### Tier of treatment/Interventions

- Adjust DRF fields to more appropriately capture appointments in a suitable format for Tier 4 treatment. This could involve the introduction of a new table related solely to residential stays, minimising impact on Tier 3 provider reporting and allowing residential stays and individual treatment appointments to be accurately reported.
- Record the PGSI and Core-10 (and other future outcome measure) scores at the first and last attended appointments against referrals. This could be populated from appointment data where applicable so no additional burden is placed on Tier 3 providers. This would enable

more flexibility in how measures can be added to the DRF for Tier 4 treatment and would make quick analysis of referral outcomes simpler.

- Review the way that service users are referred and assessed for Tier 2 and Tier 3 treatment. This may help to establish why there is such a high proportion of service users who receive only 3 or fewer appointments in a completed Tier 3 treatment journey.
- Collect interventions at the referral level at treatment end instead of the appointment level. Collecting interventions at appointment level provides the most granular data but is currently not providing usable data. Treatment end here should include service users who drop out.
- In addition, information on whether group/couples work was undertaken during the treatment referral could also be collected at treatment end.

### Other data considerations

- The DRF should provide clearer guidance around when a referral is considered “new” and when it is an existing referral restarting.
- GambleAware should work with providers to agree a consistent approach to aftercare and treatment follow-ups with the aim of improving data on continuing contact after treatment.
- Review data on affected others to ensure that it is fit for purpose, making necessary adjustments to ensure that details on the relationship to the gambler are available. Consideration should also be given to capturing when an affected other and gambler attend treatment together. Currently this would be partly captured by “Treatment Attendees” which does not explicitly identify if it involves the affected other and gambler.

# Annex 1 Missing Data Profiles

## Referral level data

The only missing values at referral level for gamblers are in service user age and gender, as shown in Table 67. Little's MCAR test (Little, 1988) had a p-value of 1, which suggests there was no strong evidence of patterns in the missing values.

***Table 67: Missing values profile for referral level variables for gamblers***

| | N | Mean | Std. Dev | Missing | |
|---|---|---|---|---|---|
| | | | | Count | % |
| Age | 12,420 | 35.6 | 10.6 | 14 | 0.1 |
| Gender | 12,361 | | | 73 | 0.6 |
| Known recurrence | 12,434 | | | 0 | 0.0 |
| Treatment completion | 12,434 | | | 0 | 0.0 |
| Total attended appts when last measure was taken | 12,434 | 6.6 | 3.6 | 0 | 0.0 |
| Total missed appointments | 12,434 | 1.7 | 1.9 | 0 | 0.0 |
| Total appointments | 12,434 | 8.3 | 4.1 | 0 | 0.0 |
| Total attended appointments | 12,434 | 6.6 | 3.6 | 0 | 0.0 |
| First PGSI Score | 12,434 | 19.0 | 6.2 | 0 | 0.0 |
| First Core-10 Score | 12,434 | 18.4 | 8.1 | 0 | 0.0 |
| Last PGSI Score | 12,434 | 4.8 | 6.1 | 0 | 0.0 |
| Last Core-10 Score | 12,434 | 7.8 | 7.6 | 0 | 0.0 |

*Little's MCAR test: Chi-Square = <.001, DF = 8, p value = 1.000*

Similar observations were made with the data for affected others, with the only missing values at referral level again being service user age and gender, as shown in Table 68. Little's MCAR test had a p-value of 0.66, suggesting that there was no strong evidence of patterns in the missing values.

*Table 68: Missing values profile for referral level variables for affected others*

| | N | Mean | Std. Dev | Missing | |
|---|---|---|---|---|---|
| | | | | Count | % |
| Age | 2,129 | 43.2 | 14.0 | 3 | 0.1 |
| Gender | 2,117 | | | 15 | 0.7 |
| Known recurrence | 2,132 | | | 0 | 0.0 |
| Treatment completion | 2,132 | | | 0 | 0.0 |
| Total attended appts when last measure was taken | 2,132 | 6.3 | 3.5 | 0 | 0.0 |
| Total missed appointments | 2,132 | 1.2 | 1.5 | 0 | 0.0 |
| Total appointments | 2,132 | 7.5 | 3.9 | 0 | 0.0 |
| Total attended appointments | 2,132 | 6.3 | 3.5 | 0 | 0.0 |
| First Core-10 Score | 2,132 | 16.6 | 7.7 | 0 | 0.0 |
| Last Core-10 Score | 2,132 | 8.0 | 7.0 | 0 | 0.0 |

*Little's MCAR test: Chi-Square = 4.14, DF = 6, p value = .66*

### Appointment level data

Table 69 shows the missing data profiles for gamblers' PGSI, while Table 70 shows the missing data profiles for gamblers' Core-10. The Core-10 had a Little's MCAR test p-value of 0.08, which provides some evidence for patterns in the missing values relating to the Core-10 outcome measure.

*Table 69: Missing data profile for appointment level data for gamblers (PGSI)*

| | N | Mean | Std. Dev | Missing | |
|---|---|---|---|---|---|
| | | | | Count | % |
| Age | 81,531 | 36.4 | 10.9 | 0 | 0.0 |
| Gender | 81,531 | | | 0 | 0.0 |
| Known recurrence | 81,531 | | | 0 | 0.0 |
| Treatment completion | 81,531 | | | 0 | 0.0 |
| Referral Number | 81,531 | | | 0 | 0.0 |
| Service user with multiple referrals | 81,531 | | | 0 | 0.0 |
| Previous appointment attendance | 69,814 | | | 11,717 | 14.4 |
| Attended appointment order | 81,531 | | | 0 | 0.0 |
| PGSI Score | 79,545 | 8.5 | 7.9 | 1,986 | 2.4 |

*Little's MCAR test: Chi-Square = 5.15, DF = 1, p value = .023*

The Little's MCAR test p-value for PGSI was 0.023, which provides strong evidence of missing data patterns. The 11,717 missing values in previous appointment attendance were due to the missing values in the first appointment only. This behaviour is expected because by definition first appointments would not have a previous appointment for which an attendance value could be

assessed. We accounted for the first appointment scores by examining change in each appointment compared to the first appointment score and then excluding first appointments from the model.

***Table 70: Missing data profile for appointment level data for gamblers (Core-10)***

| | N | Mean | Std. Dev | Missing | |
|---|---|---|---|---|---|
| | | | | Count | % |
| Age | 81,531 | 36.4 | 10.9 | 0 | 0.0 |
| Gender | 81,531 | | | 0 | 0.0 |
| Known recurrence | 81,531 | | | 0 | 0.0 |
| Treatment completion | 81,531 | | | 0 | 0.0 |
| Referral Number | 81,531 | | | 0 | 0.0 |
| Service user with multiple referrals | 81,531 | | | 0 | 0.0 |
| Previous appointment attendance | 69,814 | | | 11,717 | 14.4 |
| Attended appointment order | 81,531 | | | 0 | 0.0 |
| Core-10 Score | 79,636 | 11.6 | 8.5 | 1,895 | 2.3 |

*Little's MCAR test: Chi-Square = 3.07, DF = 1, p value = .08*

Due to missing data patterns in the PGSI score, we imputed the PGSI Scores and domains of these missing values using multiple imputation (Rubin, 1987). The multiple imputation model used in this case was linear regression as PGSI score is a continuous variable; full details of the imputation method are provided in Annex 2.

To have consistency in imputation, we also impute the missing Core-10 scores and domains using the same algorithm for multiple imputation with a linear regression model. After imputation of PGSI Score and Core-10 Score for the missingness, there are no longer any missing observations for the gambler's attended appointments.

Table 71 shows the missing data profiles for affected others' Core-10. The Core-10 had a Little's MCAR test p-value of 0.14, which suggests weak evidence of patterns in the missing values relating to the Core-10 outcome measure. However, to be consistent with our treatment of the data for gamblers, we also impute Core-10 scores and domains for the affected others using the same method.

***Table 71: Missing data profile for appointment level data for affected others (Core-10)***

| | N | Mean | Std. Dev | Missing | |
|---|---|---|---|---|---|
| | | | | Count | % |
| Age | 13,435 | 43.9 | 13.8 | 0 | 0.0 |
| Gender | 13,435 | | | 0 | 0.0 |
| Known recurrence | 13,435 | | | 0 | 0.0 |
| Treatment completion | 13,435 | | | 0 | 0.0 |
| Referral Number | 13,435 | | | 0 | 0.0 |
| Service user with multiple referrals | 13,435 | | | 0 | 0.0 |
| Previous appointment attendance | 11,390 | | | 2045 | 15.2 |
| Attended appointment order | 13,435 | | | 0 | 0.0 |
| Core-10 Score | 13,173 | 12.3 | 7.7 | 262 | 2.0 |

*Little's MCAR test: Chi-Square = 2.22, DF = 1, p value = .14*

## Annex 2 Multiple Imputation Method

Missing values for Core-10 and PGSI (and domains) where imputed for the appointment level mixed model (Table 72). Where a value is missing for one domain it will be missing for all as the provided data on Core-10 and PGSI items from which domains are derived is either missing for all items or present for all items. There are no partially missing Core-10 or PGSI item breakdowns and so no variance in the missingness of the domains from which they are calculated.

***Table 72:Variables that had missing values imputed for linear mixed model***

| Measure | Variable | With values | Missing values | |
|---|---|---|---|---|
| | | | N | % |
| PGSI | PGSI score | 79,545 | 1,986 | 2.4% |
| | PGSI domain scores | 79,416 | 2,115 | 2.6% |
| Core-10 Gamblers | Core-10 score | 79,636 | 1,895 | 2.3% |
| | Core-10 domain scores | 79,520 | 2,011 | 2.5% |
| Core-10 Affected others | Core-10 score | 13,173 | 262 | 2.0% |
| | Core-10 domain scores | 13,158 | 277 | 2.1% |

Method is taken from Rubin (1987). Let $y$ be the variable to be imputed and the predictor variables used for imputation are Age ($x_1$), Gender ($x_2$), known recurrence ($x_3$), treatment completion ($x_4$), referral number ($x_5$), Service user with multiple referrals ($x_6$), previous appointment attendance ($x_7$), attended appointment order ($x_8$), score in the first appointment ($x_9$), and the score in the previous appointment ($x_{10}$). We assume a linear regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{10} x_{10i} + \epsilon_i$$

where $\epsilon_i$ is a random error term having $N(0, \sigma^2)$ distribution of mean zero and variance $\sigma^2$. Non-informative priors are used for the parameters $\beta_0, \beta_1, \ldots, \beta_{10}$ and $\sigma^2$. We obtain the parameter estimates $\hat{\beta}_{0c}, \hat{\beta}_{1c}, \ldots, \hat{\beta}_{10c}$, and $\hat{\sigma}_c^2$ with the complete data (non-missing) using least squares, then the posterior distributions for the parameters will be multivariate normal and that for $\sigma^2$ will be an inverse chi-squared distribution, which depend on the parameter estimates $\hat{\beta}_{0c}, \hat{\beta}_{1c}, \ldots, \hat{\beta}_{10c}$, and $\hat{\sigma}_c^2$. For imputation, we first draw a sample $\beta_0^*, \ldots, \beta_{10}^*$ from the posterior distribution of the ’s and draw $(\sigma^*)^2$ from the posterior distribution of $\sigma^2$. We then generate $z_i$ from  and finally, the -th imputed value of a missing observation $y_i$ would be

$$y_i = \beta_0^* + \beta_1^* x_{1i} + \beta_2^* x_{2i} + \cdots + \beta_{10}^* x_{10i} + \sigma^* z_i$$

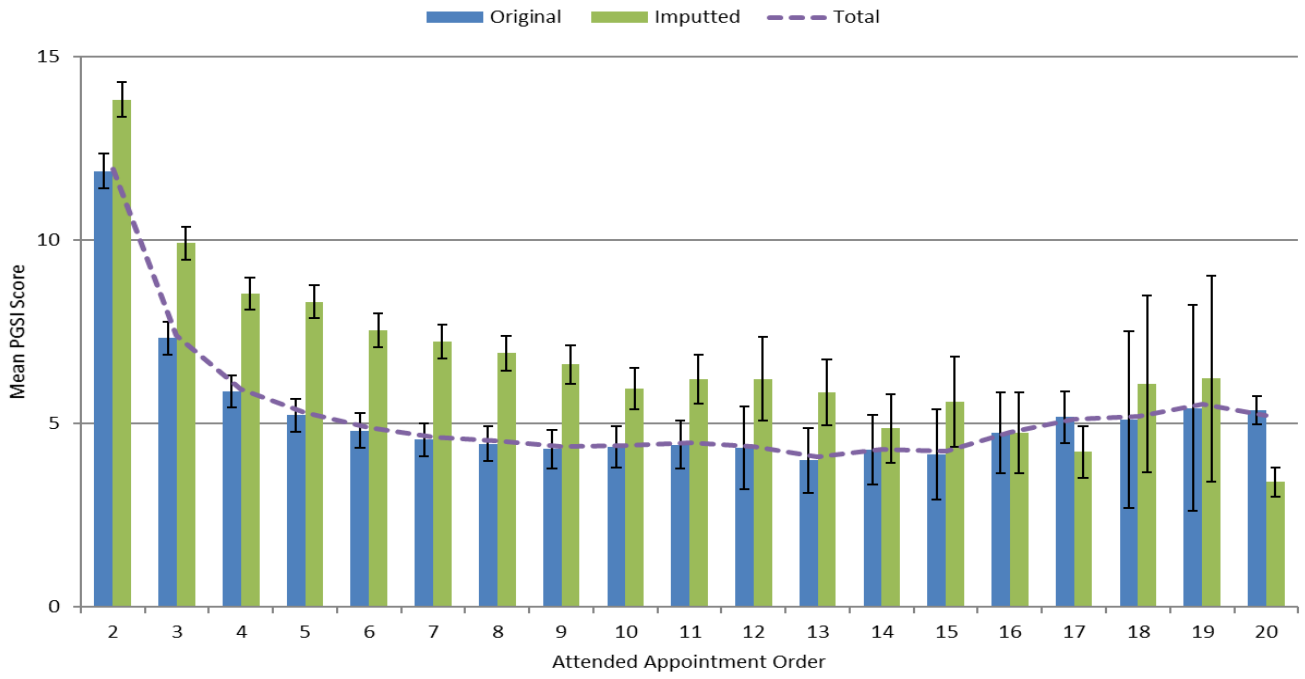
In this way, we impute all the missing values. We also incorporate restriction on the simulated value of y as part of the simulation as we know that these measurements will always be between 0 and

the maximum for the scale concerned. The imputed values are also restricted to be integers (whole numbers) as the original scores must be integers.

In general, one can have multiple imputed data. Large number of imputed datasets will reduce the variability due to imputation. We have constructed 10 imputed data sets using the method described and mean averaged (and rounded to nearest integers) the imputed data over these datasets. All mixed model analysis are based on this aggregated imputed data. If we construct a very large number of imputed data sets and aggregate then we lose out the variability in individual observation captured by the parameter σ^2 above and Rubin (1987) advised against this. One can also construct separate mixed model estimates from each of the imputed data and then pool the estimates together. However, the pooling of mixed models over multiple imputed data is not very well developed in the literature and it does not produce reliable estimates of the random effects. For this reason, we used the aggregated imputed data over 10 imputations.

As shown in Figure 19, mean imputed PGSI scores at each attended appointment order were typically slightly higher than original data by around 2 points but had little impact on the total mean at each appointment due to the low numbers involved. The main reason the imputed data is slightly higher is that mean PGSI scores in the appointment preceding the one with missing data are on average higher than for appointments with data.

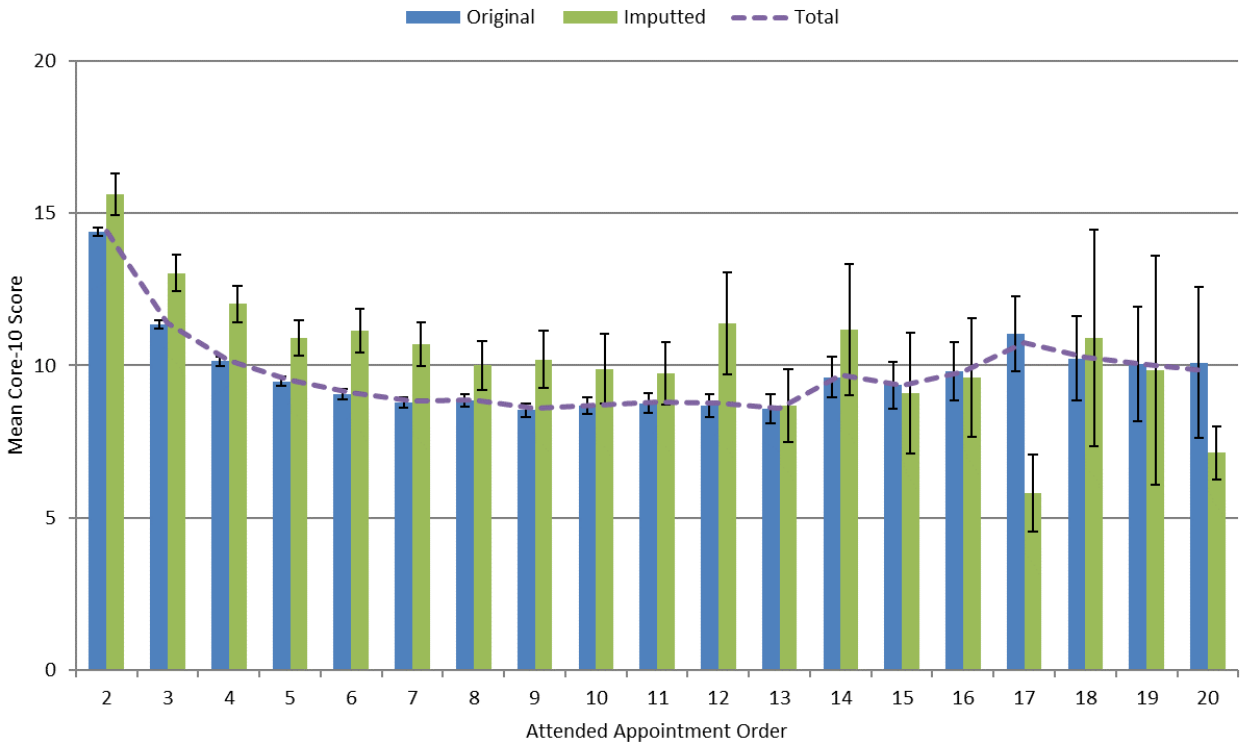
***Figure 19: Comparison of mean PGSI score (with 95% CI bars) by attended appointment order for original and imputed data***




As shown in Figure 19, mean imputed Core-10 scores at each attended appointment order were typically slightly higher than original data by around 1 points but again had little impact on the total

mean at each appointment due to the low numbers involved. The main reason the imputed data is slightly higher is that mean Core-10 scores in the appointment preceding the one with missing data are on average higher than for appointments with data, as are the mean starting scores for missing data.

***Figure 20: Comparison of mean PGSI score (with 95% CI bars) by attended appointment order for original and imputed data***



# Annex 3 Mixed model definition

We used linear mixed model analysis multi-variable regression to assess the association between attended treatment appointments on change to PGSI and Core-10 scores and domains (dependent variables listed in Table 73) from their measurements in the first appointment. This analysis was conducted separately for gamblers and affected others.

***Table 73: Dependent variables for linear mixed model***

| Measure | Variable |
|---|---|
| PGSI | PGSI score |
| | PGSI behaviour domain (1,2,3,4) score |
| | PGSI recognition and guilt (5,9) score |
| | PGSI external consequences (6,7,8) score |
| Core-10 | Core-10 score |
| | Core-10 functioning domain (2,3,4) score |
| | Core-10 depression/anxiety domain (1,5,7,8,9,10) score |

The dependant variable measure $y_{ij}$ for the $i$-th service user id and $j$-th observation for that service user is defined as the difference between the measure as listed in Table 73 for the first appointment in that referral and the measure at the appointment for the j-th observation. So it measures the difference in the measure compared to the first appointment. For example, if the measure is PGSI score, then $y_{ij}$ = PGSI score at the first appointment for the $i$-th service user for that referral PGSI score at the $j$-th appointment for the $i$th service user for that referral.

Attended appointment order was calculated using the attendance values of each appointment within a referral and ordering them according to date. For appointments that occurred on the same day but had different purposes and measures, we treated assessments as having occurred before treatment, and treatment as occurring before reviews.

The model used in this analysis is the following linear mixed model:

$$\begin{aligned} y_{ij} = {} & \beta_0 + \beta_1 \times \mathrm{Gender}_i + \beta_2 \times \mathrm{Age}_i + \beta_3 \times \mathrm{Recurrence}_i + \beta_4 \times \mathrm{CompletedTreatment}_i \\ & + \beta_5 \times \mathrm{ApptOrder}_{ij} + \beta_6 \times \mathrm{ApptOrder}_{ij}^2 + \beta_7 \times \mathrm{PreviousApptAttended}_{ij} \\ & + \beta_8 \times \mathrm{service\ userwithMultiEpi}_i + \beta_9 \times \mathrm{ReferralNumber}_{ij} \\ & + \left(\beta_{0i} + \beta_{5i} \times \mathrm{ApptOrder}_{ij} + \beta_{6i} \times \mathrm{ApptOrder}_{ij}^2 + \beta_{7i} \times \mathrm{PreviousApptAttended}_{ij} \right. \\ & \left. + \beta_{9i} \times \mathrm{ReferralNumber}_{ij}\right) + \epsilon_{ij} \end{aligned}$$

$i$ – denotes the service user id and $j$ – denotes the observation number within an individual referral.

$\epsilon_{ij}$ is the random error term having $N(0, \sigma^2)$ distribution with mean zero and variance [sigma squared].

Gender, age, treatment recurrence, completed treatment, attended appointment order, squared attended appointment order, previous appointment attendance, whether service user had multiple referrals in the data and the order of their referrals are all fixed effects with $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$ and $\beta_9$ are fixed effect coefficients.

An intercept, attended appointment order, squared attended appointment order, previous appointment attendance and the order of the service user's referrals are treated as random effects within each referral.

The coefficients $\beta_{0i}, \beta_{5i}, \beta_{6i}, \beta_{7i}$, and $\beta_{9i}$ of the random effect terms are assumed to have $N(0, \sigma_0^2)$, $N(0, \sigma_a^2)$, $N(0, \sigma_b^2)$, $N(0, \sigma_c^2)$ and $N(0, \sigma_d^2)$ distributions, respectively.

This model assumes not only a quadratic effect for the appointments but also that the quadratic curve varies randomly over the individuals. We further assume a *variance component* model, that is, $\beta_{0i}, \beta_{5i}, \beta_{6i}, \beta_{7i}$ and $\beta_{9i}$ are mutually independent and independent of the error term $\epsilon_{ij}$.

From the fixed effects, we have

$$\begin{aligned} E(y_{ij}) = \beta_0 &+ \beta_1 \times \text{Gender}_i + \beta_2 \times \text{Age}_i + \beta_3 \times \text{Recurrence}_i + \beta_4 \times \text{CompletedTreatment}_i \\ &+ \beta_5 \times \text{ApptOrder}_{ij} + \beta_6 \times \text{ApptOrder}_{ij}^2 + \beta_7 \times \text{PreviousApptAttended}_{ij} \\ &+ \beta_8 \times \text{service userwithMultiEpi}_i + \beta_9 \times \text{ReferralNumber}_{ij} \end{aligned}$$

And

$$Var(y_{ij}) = \sigma^2 + \sigma_0^2 + \sigma_a^2 + \sigma_b^2 + \sigma_c^2 + \sigma_d^2$$

$$Cov(y_{ij}, y_{ik}) = \sigma_0^2 + \sigma_a^2 + \sigma_b^2 + \sigma_c^2 + \sigma_d^2$$

We use a REML (Restricted Maximum Likelihood) algorithm to obtain the parameter estimates, the estimates of variance and covariance of the random effects terms.

### Maximum change in score

As described above, the expected change in score, $E(y_{ij})$ is given by:

$$\begin{aligned} (y_{ij}) = \beta_0 &+ \beta_1 \times \text{Gender}_i + \beta_2 \times \text{Age}_i + \beta_3 \times \text{Recurrence}_i + \beta_4 \times \text{CompletedTreatment}_i \\ &+ \beta_5 \times \text{ApptOrder}_{ij} + \beta_6 \times \text{ApptOrder}_{ij}^2 + \beta_7 \times \text{PreviousApptAttended}_{ij} \\ &+ \beta_8 \times \text{Service userwithMultiEpi}_i + \beta_9 \times \text{ReferralNumber}_{ij} \end{aligned}$$

To find the maximum change in the expected score with respect to attended appointment order, we differentiation the above equation with respect to $\text{ApptOrder}$ to get:

$$\frac{d\,E(y)}{d(\text{ApptOrder})} = \beta_5 + 2\beta_6 \times \text{ApptOrder}$$

Equating the derivative to zero, we get

$$\frac{d\,E(y)}{d(\text{ApptOrder})} = 0 \quad \Longrightarrow \beta_5 + 2\beta_6 \times \text{ApptOrder} = 0$$

$$\Longrightarrow \text{ApptOrder} = -\frac{\beta_5}{2\beta_6}.$$

If $\beta_6 < 0$ , this solution gives the maxima for the expected score with respect to appointment order. Therefore, for fixed levels of the other predictors, the maximum change in score is attained at

$$\text{ApptOrder} = -\frac{\beta_5}{2\beta_6}$$

This is estimated from the estimated coefficients of the linear mixed model as

$$\widehat{ApptOrder} = -\frac{\hat{\beta}_5}{2\hat{\beta}_6}.$$

# Annex 4 Comparison of GDIT items and existing DRF data

***Table 74: Comparison of GDIT items and existing DRF data***

| GDIT item question | GDIT item response | Existing DRF Data | Modification needed and limitations |
|---|---|---|---|
| 1. How often do you gamble? | Never, Monthly or less, 2-4 times a month, 2-3 times a week, 4 or more times a week, Daily, Several times a day | Time spent gambling – last 30 days (days)<br>This is collected as a numeric value in days via an open text field | Question would need to be converted to a multiple choice from open text. Some values could be converted although no data would correspond to 'several times a day'. Converting to multiple choice may help to reduce data entry error and improve comparability. |
| 2. How much time do you spend gambling on a typical day? | No time,<br>Less than an hour,<br>1-2 hours, 3-4 hours,<br>5-6 hours, 7-9 hours,<br>10-24 hours | Time spent gambling – daily average (hours) | Question would need to be converted to a multiple choice from open text. All values could be converted. Converting to multiple choice may help to reduce data entry error and improve comparability. |
| 3. How much time do you spend thinking about gambling on a typical day? | | None | This would need to be added as a question. |
| 5. How often have you gambled to win back money you lost on gambling, in the past 12 months? | Never,<br>Less often than monthly,<br>Monthly,<br>Weekly,<br>Daily or almost daily | PGSI 3: When you gambled, did you go back another day to try to win back the money you lost? | Potential mapping of PGSI categories would be:<br>never = never, sometimes = Less often the monthly/Monthly, most of the time = Weekly, almost always = Daily or almost daily<br>The conversion is not perfect and testing would need to examine the impact of moving from 4 to 5 response categories and the difference in wording for historical comparisons between PGSI and GDIT. It is also worth noting the timeframe in the GDIT is specified as 12 months. It would be beneficial to test whether it was possible to link timeframes to treatment, e.g. should be taken as in last 12 months at start of treatment and then reference "since entering treatment" at end of treatment. Follow-ups at 6 and 12 months could reference "since ending treatment". |
| 10. How often have you gambled with larger sums to get the same feeling of excitement as before, in the past 12 months? | | PGSI 2: Have you needed to gamble with larger amounts of money to get the same feeling of excitement? | |
| 8. How often have you borrowed money or sold something to obtain money for gambling, in the past 12 months? | | PGSI 4: Have you borrowed money or sold anything to get money to gamble? | |
| 6. How often, in the past 12 months, have you gambled more than you planned (more occasions, longer time or larger sums)? | | PGSI 1: Have you bet more than you could really afford to lose? | Mappings to PGSI as above. This is a slightly differently worded question so potentially may need to be added as a new question instead. |
| 4. How often have you tried to control, cut down or stop your gambling, in the past 12 months? | | None | This would need to be added as a question. |
| 7. How often have you lied to others about your gambling, in the past 12 months? | | None | This would need to be added as a question. |

| GDIT item question | GDIT item response | Existing DRF Data | Modification needed and limitations |
|---|---|---|---|
| 9. How often have you gambled as a way of escaping problems or relieving negative feelings, in the past 12 months? | | None | This would need to be added as a question. |
| 11. Have you or anyone close to you experienced financial problems due to your gambling? | No,<br>Yes, but not in the past year,<br>Yes in the past year | PGSI 8: Has your gambling caused any financial problems for you or your household? | Potential mapping of PGSI categories would be:<br>never = No, sometimes = Yes, but not in the past year, most of the time/almost always = Yes in the past year<br>The conversion is not perfect and testing would need to examine the impact of moving from 4 to 5 response categories and the difference in wording for historical comparisons between PGSI and GDIT. It is also worth noting the timeframe in the GDIT is specified as the past year. It would be beneficial to test whether it was possible to link timeframes to treatment, e.g. should be taken as in last 12 months at start of treatment and then reference "since entering treatment" at end of treatment. Follow-ups at 6 and 12 months could reference "since ending treatment". |
| 12. Has your gambling worsened your mental health? | | PGSI 6: Has gambling caused you any health problems, including stress or anxiety? | |
| 13. Have you experienced serious problems in any important relationship because of your gambling? | | Relationship loss through gambling | This is currently collected as a binary yes/no response with the question "Have you ever lost a relationship because of gambling?" This would need to be modified to have a broader question and more granular answers but both could potential provide more insightful data, particularly with the specified timeframes in the response. |
| 14. Have you experienced serious problems at work or in school because of your gambling? | | Job loss through gambling | This is currently collected as a binary yes/no response with the question "Have you ever lost a job because of gambling?" This would need to be modified to have a broader question and more granular answers but both could potential provide more insightful data, particularly with the specified timeframes in the response. |

## Annex 5 References


Abbott, M. W., & Volberg, R. A. (2006). The measurement of adult problem and pathological gambling. International Gambling Studies, 6(2), 175 200. https://doi.org/10.1080/14459790600928678

Alho, H., Mäkelä, N., Isotalo, J., Toivonen, L., Ollikainen, J. and Castrén, S. (2022) Intranasal as needed naloxone in the treatment of gambling disorder: A randomised controlled trial. *Addictive Behaviours* 125, 107127. www.sciencedirect.com/science/article/pii/S0306460321003129

American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) (5th edn). Arlington, VA: APA.

Aragay, N., Jimenez-Murcia, S., Granero, R., Fernandez-Aranda, F., Ramos-Grille, I., Cardona, S., et al. (2015). Pathological gambling: Understanding relapses and dropouts. Comprehensive Psychiatry, 57, 58–64.

Baño, M., Mestre-Bach, G., Granero, R., Fernández-Aranda, F., et al. (2021) Women and gambling disorder: Assessing dropouts and relapses in cognitive behavioral group therapy. *Addictive Behaviors* 123, 107085. www.sciencedirect.com/science/article/pii/S0306460321002707

Barkham, M., Bewick, B., Mullin, T., Gilbody, S., Connell, J., Cahill, J., Mellor-Clark, J., Richards, D., Unsworth, G. & Evans, C. (2013). The CORE-10: A short measure of psychological distress for routine use in the psychological therapies. Counselling and Psychotherapy Research. 13. 3-13. 10.1080/14733145.2012.729069.

Bhatia, Avantika & Gelso, Charles. (2017). The termination phase: Therapists' perspective on the therapeutic relationship and outcome. Psychotherapy. 54. 76-87. 10.1037/pst0000100.

Blank, L., Baxter, S., Buckley Woods, H. and Goyder, E. (2021) Interventions to reduce the public health burden of gambling-related harms: a mapping review. *Lancet Public Health* 6, e50–e63. www.thelancet.com/journals/lanpub/article/PIIS2468-2667(20)30230-9/fulltext

Brownlow, L. (2021) A review of mHealth gambling apps in Australia. Journal of Gambling Issues 47. https://pdfs.semanticscholar.org/b97d/d6338fc20bd623e4368f1654602a25cd67e7.pdf?_ga=2.237281856.1960259384.1642359379-38472175.1642359379

Caler, K., Garcia, J., & Nower, L. (2016). Assessing Problem Gambling: a Review of Classic and Specialized Measures. Current Addiction Reports, 3(4), 437-444.

Chebli, J.L., Blaszczynski, A. and Gainsbury, S.M. (2016) Internet-based interventions for addictive behaviours: A systematic review. Journal of Gambling Studies 32, 1279–1304.

Connell, J. & Barkham, M. (2007). CORE-10 User Manual, Version 1.1. CORE System Trust & CORE Information Management Systems Ltd.

Cooper-Samonini, E., Delaney, Z. and Bowden-Jones, H. (2021) Gambling disorder and NHS treatment: an overview. *Addiction* November/December. https://wchh.onlinelibrary.wiley.com/doi/full/10.1002/tre.826

Defife, J. A., Conklin, C. Z., Smith, J. M., & Poole, J. (2010). Psychotherapy appointment no-shows: Rates and reasons. Psychotherapy: Theory, Research, Practice, Training, 47(3), 413-417.

Dowling, N.A., Merkouris, S.S., Rodda, S.N., Smith, D., et al. (2021) GamblingLess: A randomised trial comparing guided and unguided internet-based gambling interventions. *Journal of Clinical Medicine* 10, 224. https://doi.org/10.3390/jcm10112224

Ferris J., Wynne H.J. (2001) *The Canadian Problem Gambling Index Final Report.* Ottawa, ON: Canadian Centre on Substance Abuse.

GambleAware (2021) Annual Statistics from the National Gambling Treatment Service (Great Britain): 1st April 2020 to 31st March 2021. London: GambleAware. www.begambleaware.org/sites/default/files/2020-12/annual-stats-2019-20_0.pdf

GambleAware (2021b) Data Reporting Framework: Data Specification 2.9.3 www.begambleaware.org/sites/default/files/2021-08/DRF%20Specification_2.9.3.pdf

Gambling Research Exchange Ontario (2017a) Research snapshot: A systematic review of Internet-based interventions for addictive behaviours. [About Chebli et al. 2016] www.greo.ca/Modules/EvidenceCentre/files/Chebli%20et%20al%20(2016)_Internet%20based%20interventions%20for%20addictive%20behaviours_Final.pdf

Gambling Research Exchange Ontario GREO) (2017b) Research snapshot: The impact of a self-help manual with or without therapist guidance on gambling outcomes. [About Campos et al. 2015] www.greo.ca/Modules/EvidenceCentre/files/CAMPOS~2.PDF

Gambling Research Exchange Ontario (GREO) (2017c) A brief treatment with one appointment is as effective as longer treatments for problem gambling. [About Toneatto 2016] www.greo.ca/Modules/EvidenceCentre/files/Toneatto%20(2016)%20Single-appointment%20interventions%20for%20problem%20gambling%20may%20be%20as%20effective%20as%20longer%20treatments_Final.pdf

Gambling Research Exchange Ontario (GREO) (2018) Research snapshot: Treating problem gambling with a workbook and motivational interviews. [About Boudreault et al. 2018,

reference below]
www.greo.ca/Modules/EvidenceCentre/files/Boudreault%20et%20al%20(2018)_Efficacy%20of%20a%20self%20help%20treatment%20for%20at%20risk%20and%20pathological%20gamblers_Final.pdf

Gambling Research Exchange Ontario (GREO) (2019a) Research snapshot: A psychodynamic treatment protocol for difficult-to-treat people with problem gambling. [About Mooney et al. 2019]
www.greo.ca/Modules/EvidenceCentre/files/Mooney%20et%20al%20(2019)_The%20pilotin g%20of%20a%20brief%20relational%20psychodynamic%20protocol%20for_Final.pdf

Gambling Research Exchange Ontario (GREO) (2019b) Research snapshot: Comparing treatment outcomes of two online self-directed gambling interventions. [About Hodgins et al. 2019]
www.greo.ca/Modules/EvidenceCentre/files/Hodgins%20et%20al%20(2019)_Online%20self-directed_Final.pdf

GamCare (2021) *Treatment Data Summary.*
https://issuu.com/tgdh/docs/gamcare_annual_treatment_client_profile?fr=sYzg2YzM5OTMxMzc

Gelso, C. J., & Woodhouse, S. S. (2002). The termination of psychotherapy: What research tells us about the process of ending treatment. In G. S. Tryon (Ed.), Counselling based on process research: Applying what we know. Boston, MA: Allyn & Bacon

Goodwin, B. C., Browne, M., Rockloff, M., & Rose, J. (2017). A typical problem gambler affects six others. International Gambling Studies, 17(2), 276-289.

Gordon Moody (2021) *Annual Data Report 2020/2021.* https://gordonmoody.org.uk/wp-content/uploads/2021/08/Gordon-Moody-2020-21-Data-Report.pdf

Gunstone, B. and Gosschalk, K. (2020) *Gambling Treatment and Support: On Behalf of GambleAware*. London: YouGov. www.begambleaware.org/sites/default/files/2020-12/gambling-treatment-and-support.pdf

Harvey, P. (2016) Effectiveness and efficiency in the treatment of gambling disorder: reflections on the Dodo Bird Conjecture, Journal of Health Sciences Research & Policy 21, 3, 178–182.

Hawker, C.O., Merkouris, S.S., Youssef, G.J., Dowling, N.A. (2021) A smartphone-delivered ecological momentary intervention for problem gambling (GamblingLess: Curb Your Urge): Single-arm acceptability and feasibility trial. *Journal of Medical Internet Research* 23, 3, e257861. https://pubmed.ncbi.nlm.nih.gov/33769294

Hickman, B., Chakraborty, B. and Norrie, C. (2021) *Analysis of GamCare Treatment service users*. NIHR Policy Research Unit in Health and Social Care Workforce, The Policy Institute,

King's College London. https://kclpure.kcl.ac.uk/portal/files/148491160/GamCare_Analysis_Report_Final_2_.pdf

Hodgins, D. C., Currie, S., el-Guebaly, N., & Peden, N. (2004). Brief motivational treatment for problem gambling: A 24-month follow-up. Psychology of Addictive Behaviors, 18, 293–296.

Kalbfleisch, L., Baxter, D., McKnight, S., Sztainert, T. and Sohail, M. (2020) *Rapid Evidence Review: Effective Treatment and Support for Problem Gambling*. Report prepared for the Gambling Commission, Birmingham, UK. Guelph, ON: Gambling Research Exchange Ontario. www.greo.ca/Modules/EvidenceCentre/Details/effective-treatment-and-support-for-problem-gambling

Kourgiantakis, T., Saint-Jacques, M.C. and Tremblay, J. (2013) Problem gambling and families: A systematic review. Journal of Social Work Practice in the Addictions 13, 4, 353–372.

Jimenez-Murcia S, Tremblay J, Stinchfield R, Granero R, Fernandez-Aranda F, Mestre-Bach G, et al. The involvement of a concerned significant other in gambling disorder treatment outcome. Journal of Gambling Studies. 2017;33(3):937–53.

Lee, B. K. (2002). Well-being by choice not by chance: An integrative system-based couple treatment model for problem gambling (Final report). Guelph, Canada: Ontario Problem Gambling Research Centre.

Little, R.J.A. (1988) A Test of Missing Completely at Random for Multivariate Data with Missing Values. Journal of the American Statistical Association, 83, 1198-1202. http://dx.doi.org/10.1080/01621459.1988.10478722

Locke, G. W., Shilkret, R., Everett, J. E., & Petry, N. M. (2013). Interpersonal guilt in college student pathological gamblers. The American journal of drug and alcohol abuse, 39(1), 28–32. https://doi.org/10.3109/00952990.2012.694520

Marchica, L. and Derevensky, J.L. (2016) Examining personalized feedback interventions for gambling disorders: A systematic review. Journal of Behavioral Addictions 5, 1–10.

Miller, N. V., Currie, S. R., Hodgins, D. C., & Casey, D. (2013). Validation of the problem gambling severity index using confirmatory factor analysis and rasch modelling. International journal of methods in psychiatric research, 22(3), 245–255. https://doi.org/10.1002/mpr.1392

Molander O, Wennberg P, Berman AH. The Gambling Disorders Identification Test (GDIT): Psychometric Evaluation of a New Comprehensive Measure for Gambling Disorder and Problem Gambling. Assessment. 2021 Oct 7:10731911211046045. doi: 10.1177/10731911211046045. Epub ahead of print. PMID: 34617456.

Nash, E. K., MacAndrews, N., & Bradford Edwards, S. (2018). Out of luck: An exploration of the causes and impacts of problem gambling. Citizens Advice, UK.

National Institute for Health and Care Excellence (2022) Gambling: Identification, diagnosis and management. www.nice.org.uk/guidance/indevelopment/gid-ng10210

NHS England (2019), The NHS Long Term Plan, NHS England, London, www.longtermplan.nhs.uk/wp-content/uploads/2019/01/nhs-long-term-plan.pdf.

Nilsson A, Magnusson K, Carlbring P, Andersson G, Hellner C. Behavioral couples therapy versus cognitive behavioral therapy for problem gambling: a randomized controlled trial. Addiction. 2020 Jul;115(7):1330-1342. doi: 10.1111/add.14900. Epub 2020 Jan 8. PMID: 31746075; PMCID: PMC7317433.

North, R.J., Holahan, C.J., Carlson, C.L. et al. From Failure to Flourishing: The Roles of Acceptance and Goal Reengagement. J Adult Dev 21, 239–250 (2014). https://doi.org/10.1007/s10804-014-9195-9

Orford, J., Wardle, H., Griffiths, M., Sproston, K. & Erens, B. (2010). PGSI and DSM-IV in the 2007 British Gambling Prevalence Survey: reliability, item response, factor structure and inter-scale agreement, International Gambling Studies, 10:1, 31-44, DOI: 10.1080/14459790903567132

Orford, J., Cousins, J., Smith, N. and Bowden-Jones H. (2017) Stress, strain, coping and social support for affected family members attending the National Problem Gambling Clinic, London. International Gambling Studies 17, 2, 259–275.

Parker, J.D.A. and Bauermann, T.M. (2015) Outcome of Psychological Treatments of Pathological Gambling: A Review and Meta-Analysis. Guelph, ON: Gambling Research Exchange Ontario.

Penfold, K.L. and Ogden, J. (2021) Exploring the experience of Gamblers Anonymous meetings during Covid-19: a qualitative study. *Current Psychology* [epub ahead of print] www.ncbi.nlm.nih.gov/pmc/articles/PMC8367765

Petry, N. M., & Hodgins, D. C. (2018). Commentary on Pickering et al. (2018): Problems with measurement consistency—Unique to gambling or widespread in addictions? Addiction, 113(3), 427–428. https://doi.org/10.1111/add.14121

Pfund, R.A., Whelan, J.P., Peter, S.C., Meyers, A.W. (2018) Can a motivational letter increase attendance to psychological treatment for gambling disorder? Psychological Services.

Public Health England. (2021). Gambling-related harms evidence review: Quantitative analysis of gambling involvement and gambling-related harms among the general population in England. London, UK: Public Health England. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1020883/Gambling_evidence_review_quantitative_report.pdf

Quilty, L.C., Wardell, J.D., Thiruchselvam, T., Keough, M.T. and Hendershot, C.S. (2019) Brief interventions for problem gambling: A meta-analysis. *PLOS ONE*. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0214502

Reilly, C. (ed.) (2017) *Gambling and Public Health: A Guide for Policymakers*. Beverly, MA: National Center for Responsible Gaming.

Ribeiro, E.O., Afonso, N.H. and Morgado, P. (2021) Non-pharmacological treatment of gambling disorder: a systematic review of randomized controlled trials. *BMC Psychiatry* 21, 105. https://bmcpsychiatry.biomedcentral.com/articles/10.1186/s12888-021-03097-2

Robson E. , Edwards J. Gambling decisions: An early intervention program for problem gamblers Journal of Gambling Studies, 18 (3) (2003), pp. 235-255

Roe D. The timing of psychodynamically oriented psychotherapy termination and its relation to reasons for termination, feelings about termination, and satisfaction with therapy. J Am Acad Psychoanal Dyn Psychiatry. 2007 Fall;35(3):443-53. doi: 10.1521/jaap.2007.35.3.443. PMID: 17907912.

Ronzitti, S., Soldini, E., Smith, N. et al. Gambling Disorder: Exploring Pre-treatment and In-treatment Dropout Predictors. A UK Study. J Gambl Stud 33, 1277–1292 (2017). https://doi.org/10.1007/s10899-017-9686-2

Rubin, R.B. (1987). Multiple Imputation for Nonresponse in surveys. New York: John Wiley

Sagoe, D., Griffiths, M.D., Erevik, E.K., Høyland, T., et al. (2021) Internet-based treatment of gambling problems: A systematic review and meta-analysis of randomized controlled trials. *Journal of Behavioral Addictions* 10, 3, 546–565. https://akjournals.com/view/journals/2006/10/3/article-p546.xml

Samuelsson, E., Wennberg, P., & Sundqvist, K. (2019). Gamblers' (mis)interpretations of Problem Gambling Severity Index items: Ambiguities in qualitative accounts from the Swedish Longitudinal Gambling Study. Nordic Studies on Alcohol and Drugs, 36(2), 140–160. https://doi.org/10.1177/1455072519829407

Sayre, S., Schmitz, J., Stotts, A., Averill, P., Rhoades, H., Grabowski, J. Determining predictors of attrition in an outpatient substance abuse program American Journal of Drug and Alcohol Abuse, 28 (1) (2002), pp. 55-72

Shaharabani Saidon, H., Shafran, N., & Rafaeli, E. (2018). Teach them how to say goodbye: The CMRA model for treatment endings. Journal of Psychotherapy Integration, 28(3), 385–400. https://doi.org/10.1037/int0000127

van der Maas, M., Shi, J., Elton-Marshall, T., Hodgins, D.C., et al. (2019) Internet-based interventions for problem gambling: Scoping review. *JMIR Mental Health* 6, e65.

*Walker, M., Toneatto, T., Potenza, M., Petry, N., Ladouceur, R., Hodgins, D., el-Guebaly, N., Echeburúa, E., & Blaszczynski, A. (2006). A framework for reporting outcomes in problem gambling treatment research: The Banff, Alberta Consensus. Addiction (Abingdon, England). 101. 504-11. 10.1111/j.1360-0443.2005.01341.x.*

*Walker R. (2009) Retention in treatment--indicator or illusion: an essay. Substance Use Misuse. 2009;44(1):18-27. doi: 10.1080/10826080802525967. PMID: 19137480.*

*Wardle, H., Reith, G., Best, D., McDaid, D., & Platt, S. (2018). Measuring gambling-related harms: a framework for action, Gambling Commission, Birmingham, UK, http://eprints.lse.ac.uk/89248/*

Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. Journal of Personality Assessment, 82, 50-59.

Wolpert M. (2014). Uses and abuses of patient reported outcome measures (PROMs): potential iatrogenic impact of PROMs implementation and how it can be mitigated. Administration and policy in mental health, 41(2), 141–145. https://doi.org/10.1007/s10488-013-0509-1

Woody, S. R., Detweiler-Bedell, J., Teachman, B. A., & O'Hearn, T. (2003). Treatment planning in psychotherapy: Taking the guesswork out of clinical care. Guilford Press.

World Health Organization (WHO) (2017) *The Epidemiology and Impact of Gambling Disorder and Other Gambling-Related Harm.* Discussion paper developed for the WHO Forum on Alcohol, Drugs and Addictive Behaviours, 26–28 June 2017. https://cdn.who.int/media/docs/default-source/substance-use/the-epidemiology-and-impact-of-gambling-disorder-and-other-gambling-relate-harm.pdf?sfvrsn=5901c849_2&download=true

World Health Organization (2022) *International Classification of Diseases 11th Revision* (ICD-11). [Adopted 2019. Came into effect 1 January 2022] https://icd.who.int/en

World Health Organization (2022) Addictive behaviour. www.who.int/health-topics/addictive-behaviour#tab=tab_1

Zucchella, C., Mantovani, E., Frederico, A., Lugoboni, F. and Tamburin, S. (2020) Non-invasive brain stimulation for gambling disorder: A systematic review. *FronTiers in Neuroscience* 18 August. www.fronTiersin.org/articles/10.3389/fnins.2020.00729/full