

# Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data



David M Clark, Lauren Canvin, John Green, Richard Layard, Stephen Pilling, Magdalena Janecka



## Summary

**Background** Internationally, the clinical outcomes of routine mental health services are rarely recorded or reported; however, an exception is the English Improving Access to Psychological Therapies (IAPT) service, which delivers psychological therapies recommended by the National Institute for Health and Care Excellence for depression and anxiety disorders to more than 537 000 patients in the UK each year. A session-by-session outcome monitoring system ensures that IAPT obtains symptom scores before and after treatment for 98% of patients. Service outcomes can then be reported, along with contextual information, on public websites.

**Methods** We used publicly available data to identify predictors of variability in clinical performance. Using  $\beta$  regression models, we analysed the outcome data released by National Health Service Digital and Public Health England for the 2014–15 financial year (April 1, 2014, to March 31, 2015) and developed a predictive model of reliable improvement and reliable recovery. We then tested whether these predictors were also associated with changes in service outcome between 2014–15 and 2015–16.

**Findings** Five service organisation features predicted clinical outcomes in 2014–15. Percentage of cases with a problem descriptor, number of treatment sessions, and percentage of referrals treated were positively associated with outcome. The time waited to start treatment and percentage of appointments missed were negatively associated with outcome. Additive odd ratios suggest that moving from the lowest to highest level on an organisational factor could improve service outcomes by 11–42%, dependent on the factor. Consistent with a causal model, most organisational factors also predicted between-year changes in outcome, together accounting for 33% of variance in reliable improvement and 22% for reliable recovery. Social deprivation was negatively associated with some outcomes, but the effect was partly mitigated by the organisational factors.

**Interpretation** Traditionally, efforts to improve mental health outcomes have largely focused on the development of new and more effective treatments. Our analyses show that the way psychological therapy services are implemented could be similarly important. Mental health services elsewhere in the UK and in other countries might benefit from adopting IAPT's approach to recording and publicly reporting clinical outcomes.

**Funding** Wellcome Trust.

Copyright © The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

## Introduction

In most countries, if you have a mental health problem patients cannot obtain information about the clinical outcomes achieved by the psychological therapy service that they might be considering for treatment. Additionally, when a service does hold such data, they will not usually be made public. This absence of transparency is a disservice to patients. It is also an impediment to the development of more effective health care, because it makes it difficult to study, and learn from, the variation in service outcomes.

The Improving Access to Psychological Therapies (IAPT) programme<sup>1</sup> is a rare exception to the general absence of transparency about outcomes for mental health services. Starting in 2008, the UK Government developed a plan to expand access to evidence-based psychological therapies for depression and anxiety disorders by training a new workforce and deploying it in specialised services throughout England. Every local health area (otherwise

known as a clinical commissioning group; CCG) now has an IAPT service, which provides psychological treatments in line with the stepped-care clinical guidelines<sup>2,3</sup> issued by the UK's National Institute for Health and Care Excellence (NICE). For all anxiety disorders, cognitive behavioural therapy (CBT) is recommended. For depression, a wider range of treatments (CBT, counselling, couples therapy, interpersonal therapy, and brief psychodynamic therapy) are recommended. The latest data<sup>4</sup> show that around 950 000 people a year have an initial assessment and advice from the IAPT service, with more than 537 000 going on to have a course of therapy (defined as two or more sessions), with the predominant method being CBT.<sup>4,5</sup> A distinctive feature of IAPT is the use of an outcome monitoring system<sup>6</sup> that ensures 98% of patients have scores recorded on well-validated self-report measures of depression and anxiety at the beginning and end of treatment, with CCG level summaries of such data publicly available. This system is a great improvement on

Lancet 2018; 391: 679–86

Published Online

December 7, 2017

[http://dx.doi.org/10.1016/S0140-6736\(17\)32133-5](http://dx.doi.org/10.1016/S0140-6736(17)32133-5)

See Comment page 636

Department of Experimental Psychology, University of Oxford, Oxford, UK (Prof D M Clark DPhil, L Canvin BA Hons); The Oxford Academic Health Sciences Network, Oxford, UK (Prof D M Clark, L Canvin); Department of Clinical Health Psychology, Central and North West London National Health Service Trust, London, UK (J Green PhD); Centre for Economic Performance, London School of Economics, London, UK (R Layard MSc); Department of Clinical, Educational and Health Psychology, University College London, London, UK (Prof S Pilling PhD); and Seaver Autism Center, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA (M Janecka PhD)

Correspondence to:

Prof David M Clark, University of Oxford, Oxford, UK  
david.clark@psy.ox.ac.uk

For more on NICE see  
<https://www.nice.org.uk>

### Measures of clinical outcome

The Patient Health Questionnaire 9-item (PHQ-9) score<sup>10</sup> (clinical cutoff >9) is used to measure symptoms of depression. The Generalised Anxiety Disorder 7-item (GAD-7) score<sup>11</sup> (cutoff >7) is the default measure of anxiety but services can also use more specific measures for particular anxiety disorders.<sup>9</sup> We modelled two of the standard outcome indices included in NHS Digital's reports: the proportion of patients in a CCG who have reliably improved and the proportion who have reliably recovered. Patients are reliably improved if their scores on depression or anxiety, or both, have reduced by a reliable amount<sup>9,12</sup> (ie, more than the measurement error of the scale) and neither measure has shown a reliable increase. Patients are reliably recovered if they reliably improve and their scores on both depression and anxiety are below the clinical cutoff scores<sup>9</sup> at the end of treatment. We assumed that patients (2%) without scores after treatment had not improved or recovered. The total treated cohort is the denominator for calculation of reliable improvement. Only cases above the clinical cutoffs<sup>9</sup> for depression, anxiety, or both before treatment are included in the computation of reliable recovery. The scores used to calculate reliable improvement and recovery are the last available after baseline scores, usually from the final therapy session but occasionally from an earlier session. NHS Digital does not report how often earlier session scores are used but data from an NHS Trust that provides multiple IAPT services suggest that the use of such scores is rare (2.5%; Green J, CNWL NHS Trust, personal communication).

### Possible predictors of outcome

Six possible predictors were investigated. The first predictor was the percentage of treated patients for whom a problem descriptor (ICD-10 code)<sup>13</sup> was recorded. This variable was regarded as important because the type of treatment a patient should receive is based on ICD-10 codes.<sup>13</sup> The second was the percentage of referrals who

were treated, because this variable captures the extent to which services focus on treatment, as opposed to providing only assessment, advice, and signposting. The third was the percentage of missed appointments (patient did not attend and failed to give advance warning). The fourth was the number of days patients waited between referral and starting treatment, the fifth was the number of treatment sessions, and the sixth was the Index of Multiple Deprivation (IMD),<sup>14</sup> for the area. The IMD covers seven domains, including income, employment, barriers to services, and crime. We did not use a seventh possible predictor (number of NICE-recommended depression treatments that the service could deliver) in the main analyses because preliminary analysis showed the NICE predictor did not relate to outcome. For the third and fifth predictors, in 2014–15, data were available only for the last 3 months of the study. For other predictors, data covering the full periods were used. The IMD values are from September, 2015, and represent the latest available figures.

### Statistical analysis

Analyses of predictors of IAPT performance per CCG during single years were done in R software (version 3.2.2) with  $\beta$  regression<sup>15</sup> (betareg package),<sup>16</sup> as recommended for models with continuous but bounded outcomes (eg, proportions, with all data points falling between 0 and 1).<sup>16</sup> The method exploits the flexibility of  $\beta$  distribution, which accommodates data heteroscedasticity (unequal variance) and absence of symmetry around the mean, both common features of proportions data that are less well dealt with in commonly used logistic and linear regression models. A logit link function in  $\beta$  regression renders the coefficients interpretable in terms of odds ratios (ORs). Analyses were initially run on the 2014–15 data and then repeated on the 2015–16 data.

To explore the associations between predictor and outcome variables, we first fitted simple regressions. Then we fitted a multiple regression model to each of the

	2014-15			2015-16		
	CCG (n)	Mean (SD)	Median (range)	CCG (n)	Mean (SD)	Median (range)
Reliable improvement (%)	211	60.60% (7.80)	61.50% (24.80-76.80)	209	62.47% (6.63)	62.50% (35.40-80.10)
Reliable recovery (%)	211	42.87% (7.50)	43.00% (17.60-64.60)	209	44.44% (6.34)	44.60% (20.40-58.70)
Patients finishing treatment (n)	211	2217 (1424)	1830 (335-10 470)	209	2567 (1579)	2220 (510-10 430)
Patients finishing treatment classified as a clinical case at pretreatment (n)	211	1978 (1289)	1615 (270-9650)	209	2330 (1436)	2020 (465-9635)
Patients with problem descriptor completeness (%)	211	67.75% (29.87)	75.80% (0-100)	209	76.17% (25.66)	85.07% (0-100)
Patients who enter treatment and receive a course of treatment (%)	211	58.28% (12.94)	58.32% (19.44-90.84)	209	57.74% (12.87)	57.35% (26.54-88.29)
Missed appointments (%)	208	11.30% (4.10)	10.36% (4.48-27.27)	209	11.87% (4.17)	10.87% (3.65-25.62)
Treatment appointments (n)	208	6.35 (0.98)	6.33 (4.23-8.83)	209	6.41 (0.91)	6.34 (3.99-8.62)
Days before entering treatment	211	33.73 (20.56)	28.10 (6.70-124.10)	209	30.98 (23.26)	23.60 (5.90-139.30)
Index of Multiple Deprivation	208	21.96 (8.71)	20.65 (5.80-47.40)	208	21.96 (8.71)	20.65 (5.80-47.40)

CCG=clinical commissioning group.

Table 1: Descriptive statistics for the CCG sample in 2014-15 and 2015-16

### Validation procedures

To mitigate against any bias due to excessive statistical power or other data artifacts, we repeated all analyses in permuted datasets—ie, when the associations between outcomes and predictors were disrupted by random sampling. To further investigate whether associations identified with our preferred method of  $\beta$  regression were robust, we repeated the analyses using standard logistic regression. Both validation analyses suggest our results arise from true associations between outcomes and predictors, rather than any analytical artifacts (appendix).

### Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

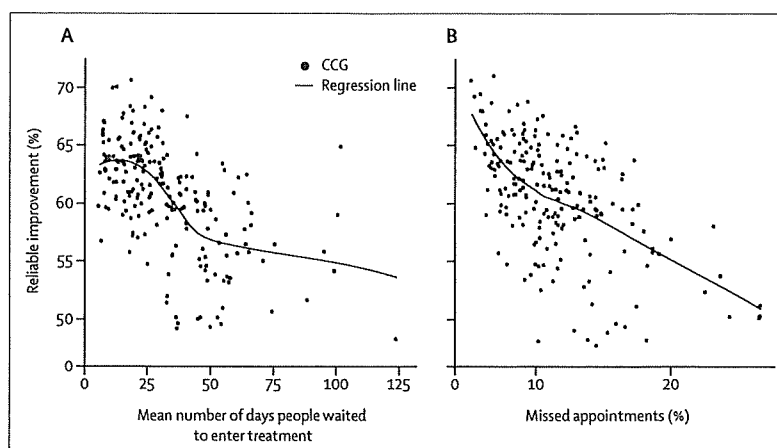
### Results

In 2014–15, 211 CCG-based IAPT services (table 1) treated 468 881 patients.<sup>5</sup> In 2015–16, mergers of some neighbouring CCGs resulted in 209 CCG-based IAPT services (table 1), which together treated 537 131 patients.<sup>4</sup>

In 2014–15, all potential predictors were significantly associated with reliable improvement and recovery when considered on their own (table 2), and for reliable recovery, all predictors were significant in multiple regression (table 2). Percentage of cases with a problem descriptor, the mean number of treatment sessions, and the percentage of referred patients who received a course of treatment were all positively associated with reliable recovery rates (table 2). Mean waiting time to enter treatment, the percentage of appointments that were missed, and the social deprivation of a CCG were negatively associated with reliable recovery rates (table 2). Mean waiting time to enter treatment, percentage of appointments missed, and percentage of referred patients who received a course of treatment were also significant predictors of reliable improvement in the multiple regression model (figure; other plots are shown in the appendix).

Table 2 shows additive ORs that estimate the changes in reliable improvement and recovery that might be achievable if the CCG with the lowest score for a particular predictor were to match the highest-scoring CCG. The analysis reveals considerable potential for improvement in the outcomes of some IAPT services. For example, if the IAPT service with the lowest percentage of patients being offered a course of treatment were to increase the proportion of people having treatment to the level of the most treatment-oriented service, recovery rates would increase by 33% and reliable improvement rates would increase by 90% (table 2).

All six predictors were individually associated with change in reliable improvement and recovery between 2014–15 and 2015–16 (appendix). In a multiple regression



**Figure:** Reliable improvement of patients in relation to time waited to start treatment in a CCG (A) and missed appointments (B)

Each blue dot represents predicted percentage of patients who reliably improve in a particular CCG, based on the predictor values recorded in that CCG and the effect sizes derived from the  $\beta$  regression. Therefore, although each graph models expected values in reference to only one of the predictors, the plotted results are a function of all six of them, accounting for the non-linear trends. Red lines were smoothed with the LOESS method, and the dark grey areas around them represent the SE around these line estimates. CCG=clinical commissioning group.

model, CCGs with the largest decreases in missed appointments and waiting times, and with the largest increases in problem descriptor completeness and percentage of referred patients who had a course of therapy, were also the CCGs that showed the largest increases in reliable improvement rates (table 3). Similar associations were noted for changes in reliable recovery, although fewer predictors were significant (table 3). Although social deprivation scores did not change between 2014–15 and 2015–16, CCGs with low amounts of social deprivation showed larger improvements in reliable recovery rates between 2014–15 and 2015–16 than CCGs with high amounts of social deprivation (table 3). Overall, the predictors we investigated explained a variance of 33% for change in reliable improvement and 22% for recovery rates.

As in 2014–15, all six predictors in 2015–16 were significantly associated with both reliable improvement and reliable recovery in single regression models.  $\beta$  regressions with multiple predictors showed that most predictors of CCGs IAPT outcomes in 2014–15 continued to be significant predictors in 2015–16 (table 4).

The 2015–16 dataset includes some potential predictors that were not available in 2014–15. Stepped care is a key organising feature of IAPT. Simple regressions showed that CCGs in which a larger proportion of patients had low-intensity treatment only had low reliable improvement and recovery rates, whereas those CCGs with high proportions of patients who had both low-intensity and high-intensity interventions (stepped care) in their course of treatment had the highest reliable improvement and recovery rates (appendix). The proportion of patients with low-intensity and high-intensity treatment was also a significant additional

NICE-recommended psychological therapies<sup>23</sup> for depression and anxiety disorders using a stepped-care model with therapists who have been through a national training programme.<sup>1</sup> Although the overall outcomes are broadly in line with expectation from clinical trials,<sup>19</sup> the performance of individual services varies greatly.

Key aspects of the way a service is organised predict a substantial amount of this variability, both within year and when modelling changes between years. The organisational factors identified make clinical sense. Services that are better able to identify the problems they are treating (problem descriptor completeness) are presumably more likely to give the right NICE-recommended treatment. Short waits between referral and treatment might ensure patients remain enthusiastic about engaging with treatment. A high average dose of treatment (number of sessions) is likely to improve outcomes, as is consistency of attendance (low percentage of missed appointments), and a service that is predominantly focused on treatment, rather than only assessment (high percentage of patients entering treatment). Focusing on treatment is reminiscent of the positive association in surgery between postoperative outcome and the volume of operations undertaken by hospitals and individual surgeons.<sup>20</sup>

The range of IAPT data that is placed in the public domain on NHS Digital and Public Health England's websites is constantly expanding. Over time it will enable interested parties to explore the importance of new variables, and study how the effect of previously studied variables changes as services work to improve the way they are organised. For the first time, NHS Digital's latest annual report<sup>4</sup> (October, 2016) includes CCG-level data on how stepped care is implemented. When we included this information in our model we found that services in which a high proportion of patients had some high-intensity therapy in their course of treatment had significantly better outcomes than those with a low proportion of patients. This outcome suggests services should make full use of stepped care, with patients who do not recover after low-intensity intervention (such as guided self-help) being given the opportunity to be stepped up to high-intensity intervention (traditional face-to-face therapy) rather than simply be discharged. A similar finding emerged from analysis of data from the first year of the IAPT programme, when only 32 services were available.<sup>21</sup> Future IAPT reports are likely to include funding information, which could be included in statistical models.

In line with previous research,<sup>22</sup> we found that social deprivation is a significant predictor of outcome. Although acting on social deprivation is a matter for local and national policy and economic development, some of its effects on outcome could possibly be mitigated by ensuring that IAPT services in socially deprived areas are of high quality and adequately funded. Consistent with this, we found that the effect sizes for social deprivation as a predictor of outcome were about

halved in the multiple regressions compared with the single regressions.

IAPT is a rare example of consistent outcome monitoring and reporting in routine mental health services. Initially, many practitioners were sceptical about the feasibility of obtaining outcome data from almost all patients treated by IAPT services. This scepticism was understandable because previous attempts that were based on data collection at only the start and end of treatment were associated with high rates of missing data for after treatment.<sup>7</sup> Furthermore, evidence that people who failed to provide data after treatment tended to have a poorer clinical outcome was worrying.<sup>23</sup> IAPT manages to obtain an end of treatment symptom measure for almost everyone (98%) who has a course of treatment by the simple manoeuvre of asking patients to complete symptom measures every time they are seen.

Now that the possibility to obtain more or less complete outcome data is clear, and that these data can be used to identify potential ways to improve patients' outcomes, we hope mental health services elsewhere in the UK and other countries will consider adopting a similar approach to data collection and reporting. In England the IAPT dataset has greatly improved public transparency about the mental health outcomes associated with routinely delivered psychological therapies. Patients can see what their local service offers and the outcomes it achieves. Commissioners and clinicians working in the services can now benchmark their service against others and consider the development of collaborative networks in which services come together to discuss common problems and learn from each other's solutions.

Our study has some limitations. Because patient-level data were not available, we could not simultaneously estimate the effects on outcome of both patient-level variables (such as initial symptom severity, sex, age, and anxiety measures) and service organisation factors. However, we were able to show that the severity of clinical problems that services treat varied very little and that this omission cannot explain our findings. Outcome assessment was restricted to patient self-report. Although our analyses illustrate potential gains achievable in the worst-performing CCGs by bringing their organisational characteristics to the level shown by the best-performing ones, we acknowledge that these are predictions that will need to be further tested.

Before IAPT, little was known about the outcomes achieved in routine mental health services. The session-by-session outcome monitoring system used in IAPT enables the services to collect outcome data from almost everyone. Publication of this data is improving public transparency and also allows analyses to be done that help us to understand, and hopefully reduce, local variability in mental health outcomes.

#### Contributors

DMC contributed to the study design, data acquisition, analysis plan, literature search, drafting of the manuscript and appendix, and final