Community Mental Health Journal (2019) 55:873–883
https://doi.org/10.1007/s10597-019-00390-z

ORIGINAL PAPER

The Cost-Effectiveness of the Improving Access to Psychological Therapies (IAPT) Programme in Severe Mental Illness: A Decision Analytical Model Using Routine Data

Darshan Zala[1] · Alison Brabban[2] · Alex Stirzaker[2] · Muralikrishnan Radhakrishnan Kartha[1] · Paul McCrone[1]

Received: 20 February 2018 / Accepted: 1 March 2019 / Published online: 8 March 2019

Abstract
This is the first site level economic evaluation of the Improving Access to Psychological Therapies programme for severe mental illness (IAPT-SMI) that is funded by NHS England. It also aims to illustrate the challenges involved in evaluations based on routine data with low internal validity. Six IAPT-SMI pilot sites treated 1 of 2 clinical groups: (i) psychosis or bipolar disorder; (ii) personality disorder. A decision analytical model nested in a before-after framework- the same patients 12 months after treatment versus 12 months before treatment—was used to compare the cost-effectiveness of IAPT-SMI with treatment as usual (TAU). IAPT-SMI appears to be more costly overall but save non-psychological treatment costs. There is evidence it may improve function and lower incidence of harmful behaviour. However, there is a need for evaluations with a more conventional study design that measure a more comprehensive array of resource use and clinical outcomes.

Keywords Severe mental illness · IAPT · Improving access to psychological therapies · Cost-effectiveness · Economic evaluation

Introduction

There is increasing concern in the UK about the social and economic burden of severe mental illness (SMI) which includes schizophrenia, affective psychosis (bipolar disorder) and personality disorders. The UK prevalence of these for all severities in adults 18 and over was at least 1% as of 2007 (National Centre for Social Research 2009). An older estimate suggested 6 in 1000 people have a SMI that warrants long-term care (Charlwood et al. 1999). In the UK, the estimated service and productivity costs of mental illness are between £50 and −£100 billion a year (Davies 2013; McCrone et al. 2008), with the costs of SMI alone being estimated to be around £17 billion a year (McCrone et al. 2008).

✉ Darshan Zala
darshanzala@hotmail.co.uk

[1] Department of Health Service and Population Research, King's Health Economics (KHE) Institute of Psychiatry, Psychology & Neuroscience, David Goldberg Centre, King's College London, De Crespigny Park, Denmark Hill, Box 24, London SE5 8AF, UK

[2] Avon and Wiltshire Mental Health Partnership NHS Trust, Jenner House, Langley Park, Chippenham SN15 1GG, UK

Despite the burden of these diseases the treatment gap for SMIs has been estimated at 15–18% (Kohn et al. 2004).

There have been previous economic evaluations of psychological therapies for SMI and the results have been mixed. Many trial based evaluations have failed to show that cognitive behavioural therapy (CBT) for psychosis or personality disorder is more cost-effective than treatment as usual (TAU) (Davidson et al. 2010; Durham et al. 2003, 2005; Gumley et al. 2003; Palmer et al. 2006). There is some evidence CBT for schizophrenia spectrum disorder may slightly lower service costs over 2 years (Startup et al. 2005). A small trial suggested cognitive remediation therapy has better outcomes than TAU at no additional cost (Patel et al. 2010). There is evidence social recovery orientated cognitive behavioural therapy is more cost-effective than case management (Barton et al. 2009).

The Improving Access to Psychological Therapies programme (IAPT) began in 2008 in England and provides evidence based treatments to over 900,000 people with anxiety and depression each year (NHS England). In 2011 the UK government made a commitment to expand this service to people with SMI—specifically people with psychotic disorders, bipolar disorders and personality disorders. NHS England has since funded a pilot study—from November

2012 to March 2015—with 6 demonstration sites providing IAPT services to people with SMI (McPin Foundation 2015). Although there has been an economic evaluation of the original IAPT program using site level data (Mukuria et al. 2013), no such evaluation has been undertaken for IAPT-SMI.

NHS England provided us with aggregated real world data from a sample of service users at these demonstration sites. We employed decision analytical models in a before-after design to compare the same group of patients before and after IAPT treatments. The aim was to assess the cost-effectiveness of IAPT-SMI (the after group) compared to TAU (the before group). The study also illustrated the challenges of conducting economic evaluations based on routine real world data from multiple sites.

Methods

Study Design

The NHS England pilot study collected routine data from patients referred to 6 IAPT demonstration sites. The 6 urban and rural IAPT sites were Barnet, Enfield and Haringey ("BEH"), Birmingham and Solihull ("BS"), Northeast London ("NEL"), South London and Maudsley ("SLAM"), Somerset ("Som") and Lancashire early intervention service ("Lanc-EI"). In most cases we received a data return per site; however the data we received for 2 sites were divided into different services. For South London and Maudsley we received separate data returns for the early intervention service ("SLAM -EI") and promoting recovery service ("SLAM-PR"); for Somerset we received data returns for primary care ("Som-PC") and secondary care ("Som-SC"). Each IAPT-SMI pilot site treated only 1 of 2 clinical groups: (i) psychosis or bipolar disorder; (ii) personality disorder.

This sort of routine data differs from traditional randomised controlled trials or observational studies for five main reasons. First, there was no assignment to active treatment and comparator arms. Second, recruitment was ad hoc in that no strict protocol was applied and resources and processes for recruitment were likely to have varied between sites. Third, the amount of missing data differed greatly by reported outcome measure. This restricted the analysis in that not all relevant outcomes could be used in the model—for example, a measure of functional impairment was used (WSAS) but no scales for overall quality of life. Fourth, there was low internal validity and a likelihood of selection bias because no randomisation and blinding of patients occurred. However, this can also be a problem with observational studies as well. Fifth, external validity may have been lacking because the sample was not random and so may not be representative. However, a compensating advantage is that routine data will be more representative of real world clinical practice and the SMI pool of patients that can be effectively treated in real world practice (i.e. in contrast to controlled conditions). The latter is an important advantage because patients with SMI can be a particularly challenging group to keep in treatment. An additional challenge was that NHS England did not provide us with individual level data, which would have allowed the use of econometric and time to event analyses for more sophisticated modelling.

This was a 12 month evaluation and so no discounting of costs and outcomes was required. We report on 4 outcomes; WSAS and 3 harmful behaviours (deliberate self-harm, suicide attempts and instances of violence towards others). Costs were from an NHS perspective and are reported in 2015 British pounds. Probabilistic sensitivity analysis (PSA) was used to evaluate parameter uncertainty (Briggs et al. 2003).

Samples

The IAPT-SMI service was designed for service users with a diagnosis of SMI whose needs could be appropriately met within a psychological therapy service and who were deemed capable of engaging with the service (Jolley et al. 2015). Referrals of patients with SMI were accepted from primary and secondary care, with screening by clinicians to ensure referral criteria were met. The referral process reflected the nature of routine data research: "there are no clear inclusion criteria, service users present with complex, diagnostically indeterminate problems and staff work pragmatically and eclectically, using an admixture of published evidence, experience, and case by case science in practice" (Jolley et al. 2015). Accepted referrals were contacted by an independent assessor and people choosing to opt-in to the service were offered a pre-therapy assessment. Therapists offered a first therapy appointment as soon as possible after this assessment.

Each of the sites treated patients in only 1 of 2 sets of clinical groups defined by type of SMI: (1) Psychosis or bipolar disorder: BS, SLAM-EI, SLAM-PR and Lanc-EI, (2) Personality disorder: BEH, NEL, Som-PC and Som-SC. The data was site level and the data returns contained categorised data in the form of summary statistics such as counts and averages. Data returns provided these statistics for the analysis sample 12 months before the start of IAPT treatment and the 12 months following treatment, which facilitated the use of a before-after framework. In contrast to individual level data it was impossible to match individuals to characteristics, treatments, resource use and outcome measures.

The data returns only contained patient summary statistics of those that attended a session and these numbers will give some indication of the composition of patients used in our analysis. The majority of those across sites were female

(71.1%) although the SLAM-EI and Lanc-EI sites were an exception. Most attendees were aged 18–64, with very few attendees under 18 years of age or 65 and over. In total there was roughly a 50/50 split between the 18–35 and 36–64 group (49.2% and 47.3%), although numbers were skewed towards the younger group in the early intervention sites which is to be expected. The majority of attendees were white (76%) but this was not the case in BEH, SLAM-EI and SLAM-PR.

IAPT Interventions

The IAPT program is composed of a range of psychological therapies that are provided to patients within each site according to the individual needs of this complex group of people. Patients could be offered multiple interventions. These included individual therapies such as CBT for psychosis and personality disorder, cognitive analytical therapy (CAT), dialectical behaviour therapy (DBT), Mentalization-based treatment (MBT), art therapy, guided formulation, structured clinical management and psychodynamic talking therapy. Group and family interventions included DBT group, DBT skills group, emotional skills group and family intervention for psychosis.

Outcomes

Data returns had space for data on a number of clinical or quality of life scales but only the WSAS had relatively complete data from each site. Therefore WSAS was the only scale included in the cost-effectiveness analysis. It is a self-report scale of functional impairment and evidence suggests it is a valid measure with high sensitivity across a broad range of disorders (Mundt et al. 2002). Lower scores suggest an improvement in function. It was reported by every site although loss to follow-up (i.e. attrition) was relatively high in 4 data returns (BEH, SLAM-EI, SLAM-PR and Lanc-EI). Patients were classified into discrete percentage point change categories in total score from pre-treatment to post-treatment. There were 11 percentage point change categories reported: no change and worsening or improvements of 20% or less, 21–40%, 41–60%, 61–80% and 81–100%.

The sites working with people with personality disorder also reported incidence of the following harmful behaviours: deliberate self-harm, suicide attempts, instances of violence towards others. These were defined in accordance with local NHS trust guidance and in general were consistent with National Institute for Health and Care Excellence guidelines on mental health. They were reported as summary statistics covering two 12 month periods—12 months up to the first assigned IAPT treatment session and 12 months after the last treatment. All 3 of these outcomes were included in the analysis. However, only 2 data returns (BEH and NEL) contained data for both 12 month periods and so this outcome was only reported for the personality disorder group with an effective sample size of 833 (completers, n = 606; drop-outs = 227). Attrition was also higher for this outcome, suggesting that complete cases were followed up at a different time for this data.

Resource Use

Data returns reported statistics relevant to resource use for IAPT treatments, home treatment (HTT) team contacts and mental health acute admission bed days. For the analysis sample, the data returns reported total sessions for every therapy provided at the site. The statistics for HTT contacts and bed days were structured in the same way as the statistics for harmful behaviours—over 12 month periods before and after treatment. Only the Som-PC and Som-SC data returns provided inadequate data for our analysis and so the effective sample sizes were as follows: 726 in the psychosis or bipolar disorder group (completers, n = 599; drop-outs, n = 127) and 833 in the personality disorder group (completers, n = 606; drop-outs, n = 227).

Before resource use statistics could enter the model they were multiplied by unit costs. For IAPT therapy treatment costs, this costing process began by multiplying the total number of sessions of each IAPT therapy by the relevant unit cost and summing these to obtain per site total treatment costs for the analysis sample. This was then divided by the size of the analysis sample for the relevant site to obtain an average treatment cost per patient for each site.

Unit costs were obtained from published literature. The unit cost for an individual CBT session in psychosis was £117 and was an average of two published costs (Barton et al. 2009; Kuipers et al. 1998); a session of CBT for SMI that is not psychosis was costed as £98 (Curtis 2015). The unit cost for a family intervention in psychosis was £249 (Curtis 2015). A HTT visit was costed as £37 (Curtis 2015). An acute admission bed day was costed as £360 and was an average of bed day costs across all mental healthcare clusters in the 2015 NHS reference costs (codes MHCC00 to MHCC21) (NHS reference costs 2015).

Where a unit cost was not available in the literature, we attached unit costs for treatments that were most similar based on the following criteria: type of professional involved, target clinical group and the clinical nature of the therapy. MBT individual, structured clinical management, CAT and DBT were costed as CBT. Art therapy and guided formulation were costed as behavioural activation therapy (£17). DBT group and DBT skills group were costed as mindfulness group therapy (£14). Psychodynamic therapy was costed as a session with a psychiatric consultant (£139)—a clinical psychologist is often used, but this will make little difference as the cost is almost identical. These costs were for 2015 and obtained from the Personal Social Services Research Unit (Curtis 2015).

To summarise for the combined cohort, an average completer (drop-out) was costed as having the following sessions: 41 (8) individual CBT in psychosis, 51 (21) individual CBT (not psychosis), 2 (0) family interventions in psychosis, 57 (21) mindfulness group therapy, 31 (7) behavioural activation therapy and 1 (0) psychologist or psychiatrist.

Analysis

Analysis Rationale

The routine nature of the data and its structure informed the type of analysis that was conducted. The purpose of randomisation is to balance the arms in terms of factors that may confound the treatment comparison (i.e. selection bias). To compensate for this lack of internal validity, a before-after analysis was employed—by comparing the same cohort before and after a treatment is provided, we can control for various confounding patient characteristics. Similarly, the nature of a before-after design means that treatment effects are calculated in a way that makes them within site effects, which helps to control for differences between sites. This design can also help to minimise attrition bias compared to a parallel group design because many of the variables that correlate with both treatment efficacy and loss to follow-up will be balanced between arms. The structure of the data returns was also highly conducive to this approach—all 4 outcomes were recorded before and after the end of assigned treatments. Therefore, this analysis compared the total cohort in the 12 month period after the last treatment session ("IAPT arm") with the cohort 12 months before the start of the first treatment session ("TAU arm").

The aggregate nature of the data was also well suited to decision analytical modelling, particularly decision tree analysis. Moving from the left to the right of a decision tree, each chance node (circle) represents a set of mutually exclusive (and jointly exhaustive) possibilities for a hypothetical patient (Petrou and Gray 2011). All pay-offs (cost and benefit inputs) are weighted by the relevant probabilities and the sum of these will give the expected (i.e. average) total pay-offs for the patient over 12 months.

All cost and outcome data in the data returns—even in the pre-treatment (TAU) group—were separately reported by eventual drop-out or completer classifications. We could not avoid this demarcation in the decision tree models if we wished to use as much information as possible in the analysis.

First, decision trees were designed so that the summary statistics provided in site data returns could be used as inputs. These decision tree models produced outputs (i.e. expected pay-offs) by arm (IAPT-SMI and TAU) that equate to costed resource use (HTT and acute bed day) and outcomes (3 harmful events and WSAS) for the average patient in each site. The model schematics and associated formulae to produce the expected pay-off results are presented in this section. Second, incremental (i.e. IAPT-SMI vs. TAU) average results were calculated for each site. Third, these per site incremental results were combined in a weighted average to produce results for the two combined cohorts of patients: psychosis or bipolar disorder and personality disorder. Fourth, uncertainty was investigated using conventional health economics methods: PSA, deterministic sensitivity analyses and outlier analyses. Figure 1 presents the pathway by which data for an individual in the analysis dataset was used to produce cost-effectiveness results.

Decision Tree Models for Costed Resource Use and Harmful Events

The decision tree structures for costed resource use (HTT and bed days) were the same as for the 3 harmful behaviours because the data was structured in the same way (Fig. 2).

Depending on the decision tree, an "event" in the following discussion can refer to an incidence of a harmful behaviour or a resource use contact. The data returns provided information to calculate the following variables: probability of eventual completer or drop-out status; 12 month probability of having >0 mean events per month conditional on eventual completer or drop-out status; and mean events per month for those who had an event. Model calculations use the standard probability rules associated with a decision tree. The probability of each mutually exclusive route is calculated by multiplying the respective probabilities along the route; this is multiplied by the relevant pay-off; and these expected pay-offs are summed to provide the total expected pay-off per patient. After adjustment of the mean events per month to 12 month variables the expected pay-off could be calculated.

$$
\begin{aligned}
&\text{E(total pay-off)}\\
&= \text{P(completers)} * \text{P}(>0\,\text{mean events}|\text{completers})\\
&\quad * (\text{mean events for those who had an event}|\text{completers})\\
&\quad + \text{P(drop-outs)} * \text{P}(>0\,\text{mean events}|\text{drop-outs})\\
&\quad * (\text{mean events for those who had an event}|\text{drop-outs}).
\end{aligned}
$$

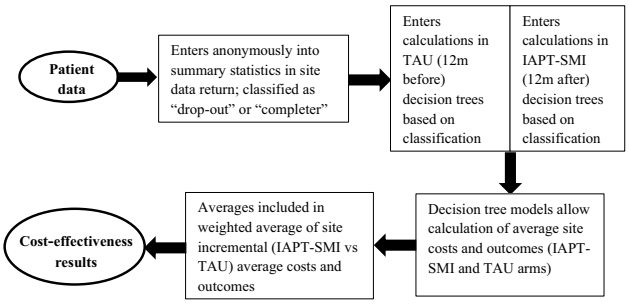

Fig. 1 From patient data to cost-effectiveness results. Schematic showing how information on a patient included in the analysis dataset is used at each stage of the analysis

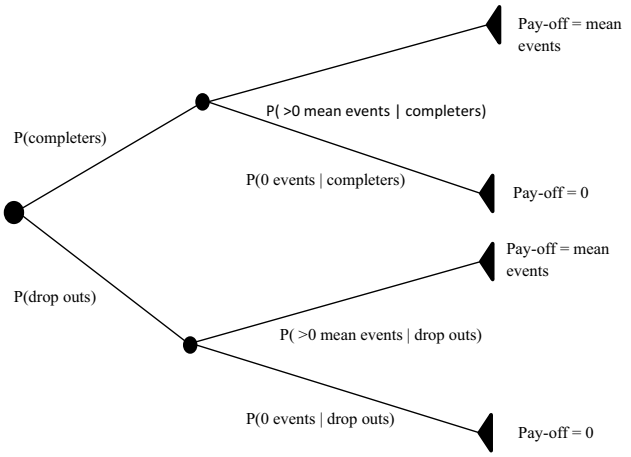

Fig. 2 Model schematic: costed resource use and the three harmful behaviours. Where data allowed, this decision tree structure was used for both arms (IAPT v TAU) in each site and for the following 5outcomes (i.e. "events"): costed HTT resource use, costed mental healthcare acute bed days and all 3 harmful events. Thehypothetical patient can either be an eventual completer or drop-out (see "Samples" section). Following this they can eitherexperience > 0 events per month or zero events. Probabilities and pay-offs will vary by outcome, site and by arm (see"Appendix 1").

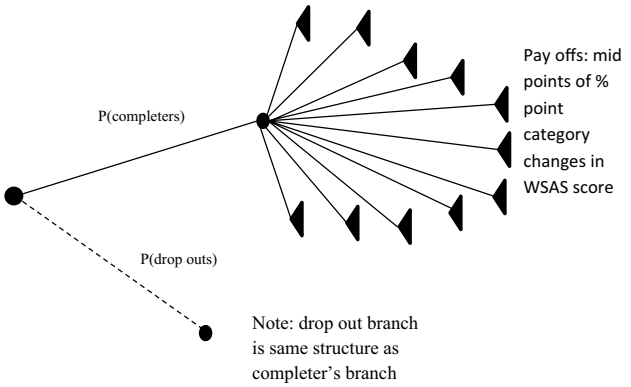

Fig. 3 Model schematic: WSAS total score. This decision tree structure was used for WSAS total score. The hypothetical patient can either be an eventual completer or drop-out (see "Samples" section). They can then fall into the following 11 categories of percentage change in WSAS score: 20% or less, 21–40%, 41–60%, 61–80% and 81–100%. Pay-offs correspond to the midpoint of each of these categories. Probabilities and pay-offs will vary by site and by arm (see "Appendix 1")

The model schematic expresses the above formula for expected pay-off as branches of a decision tree. It shows the possibility of patients experiencing zero events—however, the terms for these arms do not enter the calculation because the pay-off for them is zero. For each of the HTT and acute bed day models, events were multiplied by the relevant unit cost before being entered.

Decision Tree Models for WSAS

For the WSAS model (Fig. 3), the data returns provided information for the following variables: probability of eventual completer or drop-out status; probability of being in each category conditional on completer or drop-out status; and midpoint of each percentage point change category (ei). The "∑" indicates the summing operation across the arms representing each category after the probabilities along the route and the pay-off (i.e. mid-point of % point change category) have been multiplied.

$$\mathrm{E}(\%\text{ point change in WSAS}) = \sum[\mathrm{P}(\text{completers}) * \mathrm{P}(e_i|\ \text{completers}) * (\text{midpoint of ei})] + \sum[\mathrm{P}(\text{drop}-\text{outs}) * \mathrm{P}(e_i|\ \text{drop}-\text{outs}) * (\text{midpoint of ei})].$$

Obtaining Combined Cohort Results

All inputs for each site are shown in "Appendix 1" so that pay-off calculations can be replicated. Therefore we obtained per site average differences in total cost and outcomes for IAPT relative to TAU. A total average difference for the whole cohort was obtained by taking an average of these and weighting by the size of the analysis sample in each site. This process preserved the within cohort comparisons that help reduce selection bias—it is analogous to using a meta-analysis to combine treatment effects from different trials and is in contrast to pooling the data. Therefore only incremental results were calculated and reported. This was done separately for those sites that provided treatments for psychosis or bipolar disorder, and those that provided treatments for personality disorder. Incremental cost-effectiveness ratios (ICERs) were calculated for each of these groups and gave the average cost of providing each additional benefit on each outcome measure for IAPT relative to TAU.

Investigating Uncertainty in Cost-Effectiveness Results

Parameter uncertainty was investigated by a PSA that involved 4000 Monte Carlo simulations from distributions pre-assigned to the 401 probability and cost parameters used as inputs into the models (Briggs et al. 2003). Probabilities were assigned Beta distributions or the multivariate equivalent (Dirichlet) where appropriate; costs were assigned Gamma distributions. The PSA produced 4000 simulations of the model results, the spread of which reflected

the underlying stochastic uncertainty in inputs propagated through the model to the output. The 2.5% and 97.5% percentiles of these simulations were used to construct 95% credible intervals (CrI). There is a 95% probability that the point estimate of interest will lie in this interval - the wider the interval the more stochastic uncertainty around model output. Cost-effectiveness planes were also plotted so as to present uncertainty around our results. The percentage of simulations in each quadrant gave the probabilities that each of the following mutually exclusive realities would be realised: IAPT is costlier and more effective (on the WSAS scale) than TAU; costlier and less effective; less costly and less effective; and less costly and more effective. Structural uncertainty was investigated via a deterministic sensitivity analysis that increased and decreased the following modelling inputs by 20%: average treatment costs, all non-TX costs (HT team and acute bed days) and other event pay-offs (harmful events and WSAS % point change).

Ethics Approval

No ethical approval was required for this analysis because it does not contain any studies with human participants performed by any of the authors. The data provided by NHS England was routine anonymous summary data from services and so no such approval was required.

Results

Attrition is defined in terms of proportion of patients lost to the 12 month follow-up (IAPT arm) compared with 12 months before treatment (TAU arm). For the WSAS data, attrition varied widely (7% in BS to 86% in BEH) but for the total sample it was 23%. For resource use data total attrition was slightly higher at 38%. Data for harmful behaviours was only reported by 2 sites and attrition for this total sample was very high (75%). However, as already stated the before-after design can help to minimise the bias associated with attrition.

Incremental point estimates (Table 1) indicate that the IAPT arm on average is more costly by around £700 per patient—inclusive of treatment and non-psychological treatment costs—over 12 months for both clinical groups. However, the 95% credible intervals suggest there is a very high level of uncertainty around these point estimates. Point estimates also suggest that IAPT saves HTT and acute bed day costs, but this saving is eroded by the substantial treatment costs. The scatter of simulations (Figs. 4, 5) along the incremental cost dimension suggests 72% and 84% probabilities of IAPT having higher costs, respectively.

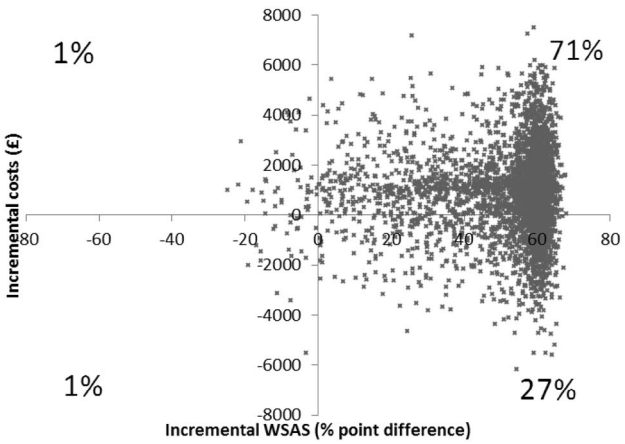

Fig. 4 WSAS cost-effectiveness plane for psychosis and bipolar disorder. The cost-effectiveness plane plots the 4000 PSA simulations. A simulation falls into a particular quadrant depending on the incremental cost (IAPT total average cost—TAU total average cost) and incremental WSAS (% point difference between IAPT and TAU) for that run of the PSA. IAPT is costlier and more effective (on the WSAS scale) than TAU in the NE quadrant; less costly and more effective in the SE; less costly and less effective in the SW; and more costly and less effective in the NW. The percentage of simulations that land in each quadrant is also shown. The WSAS scale has been reversed so that higher scores show an improvement

Table 1 Per person incremental and ICER results

Costs and outcomes	Psychosis or bipolar disorder		Personality disorder	
	Difference (95% CrI)	ICER	Difference (95% CrI)	ICER
Treatment costs (£)	£1255	–	£1634	–
HTT contact costs (£)	−£84	–	−£148	–
Acute bed days (£)	−£505	–	−£735	–
Total costs (£)	£665 (−£4763, £6604)	–	£751 (−£3004, £3856)	–
WSAS (% point)	51.5 (2.3, 89.1)	£12.9	12 (−9.9, 37.8)	£62.4
Self-harm prevented	–	–	50.4 (18.6, 71.8)	£14.9
Suicide attempt prevented	–	–	3.6 (0.9, 5)	£205.9
Violence prevented	–	–	4.8 (0.2, 6.6)	£157

CrI credible interval, *ICER* incremental cost-effectiveness ratio, *WSAS* work and social adjustment scale

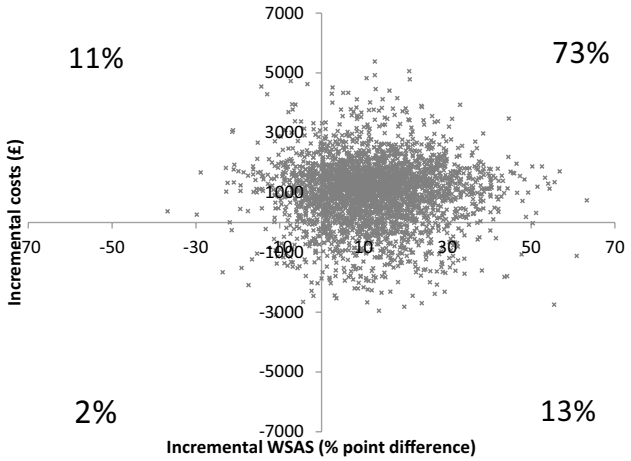

Fig. 5 WSAS cost-effectiveness plane for personality disorder. The cost-effectiveness plane plots the 4000 PSA simulations. A simulation falls into a particular quadrant depending on the incremental cost (IAPT total average cost—TAU total average cost) and incremental WSAS (% point difference between IAPT and TAU) for that run of the PSA. IAPT is costlier and more effective (on the WSAS scale) than TAU in the NE quadrant; less costly and more effective in the SE; less costly and less effective in the SW; and more costly and less effective in the NW. The percentage of simulations that land in each quadrant is also shown. The WSAS scale has been reversed so that higher scores show an improvement

Point estimates were a weighted mean of average incremental costs for each site ("Analysis" section). These were similar for most sites in the psychosis and bipolar group—an average of £1606—excluding the outlier site BS, which skewed the average down with an incremental cost of −£1960. Almost double the number of acute bed days in the TAU arm relative to the IAPT arm drove this difference in BS. The point estimate for personality disorder was partly skewed by an outlier; BEH had an average incremental cost of £1432 which was driven by high average IAPT treatment costs.

The WSAS scale has been reversed so that higher scores show an improvement in function. In the psychosis and bipolar disorder group, IAPT patients have an average WSAS score 51.5% points lower (i.e. improvement) than patients in TAU and the associated 95% credible interval does not contain 0 suggesting this estimate is less uncertain. The average score for IAPT patients is only 12% points better in the personality disorder sites but the credible interval suggests the true estimate could be 0 or less. Cost-effectiveness planes show a 98% and 86% probability of WSAS improvement, respectively. These results reflect the raw data, in terms of the way patients fall into each of the WSAS %point change categories in the 2 groups. ICERs suggest it costs an average of £13 and £62 for IAPT to improve WSAS by 1 more %point, respectively. WSAS incremental scores were stable across sites, particularly for the psychosis and bipolar group. In the personality disorder analysis, BEH was the outlier site with a % point change of −0.82, compared to an average of 13.4 for the other sites. However, this did not skew the total average a great deal after weighting by site sample sizes.

Site data would only allow an analysis of harmful behaviours for the personality disorder group. We see that for all behaviours, average per patient incidence of these behaviours is lower in the IAPT group than TAU and credible intervals suggest there is a high probability of this. The additional cost to prevent an incidence of each behaviour was estimated at £15 (self-harm), £206 (suicide attempt) and £157 (violence towards others). There were small differences in events avoided between the 2 sites that informed the point estimates for harmful behaviours. The largest difference between sites was for instances of self-harm: 41.1 for BEH and 61.1 for NEL.

Results were most sensitive to variation in treatment costs during sensitivity analyses. However, the results were robust to different scenarios—the relative positon of each arm did not change from base-case results.

Discussion

This study was a preliminary evaluation of IAPT-SMI and suggests that although IAPT is more costly overall, it saves substantial non-psychological treatment costs over 12 months—particularly acute secondary care costs. It suggests the program can improve functional impairment in patients with psychosis or bipolar disorder, and also reduce harmful behaviours in patients with personality disorders. Results were robust to deterministic sensitivity analysis, but incremental costs in particular were sensitive to outlier sites. These outlier sites may be evidence our overall results underestimate savings in acute secondary care costs, but also underestimate the costs of IAPT interventions.

Given the nature of the data used in this analysis, comparison with other research is warranted. Evaluations of individual treatments for SMI report a wide variation in incremental costs. Studies over mainly 2 years report no difference, or cost savings as high as £689 for CBT versus TAU (Durham et al. 2005; Palmer et al. 2006; Patel et al. 2010; Startup et al. 2005); another reports a saving of £6472 over 6 years (Davidson et al. 2010). Our results would not be obviously inconsistent with these if we assume that incremental non-psychological treatment cost savings continue to accrue in a similar way to the 12 months of our analysis. A study evaluating social recovery orientated cognitive behavioural therapy implied an incremental cost very similar to ours over a 9 month period (£668) (Barton et al. 2009).

A site evaluation of conventional IAPT suggested an 8 month cost difference of £163 for IAPT versus TAU

(Mukuria et al. 2013); results showed small average savings in indirect non-psychological treatment costs, but these were again dwarfed by the larger increase in treatment costs. It reported that conventional IAPT involved an average of 3 sessions of mainly CBT at an average cost of £559 per patient. Our data returns usually report greater numbers of therapy sessions and the mean of average site treatment costs was £1492, which is to be expected in the treatment of SMI.

WSAS was reported as percentage difference between IAPT and TAU, which is unusual. For a number of mainly CBT trials we calculated WSAS total score at follow-up in the active treatment arm as a percentage of the score at follow-up in TAU (Murphy et al. 2015; Ridsdale et al. 2012; Talbot et al. 2014; Zu et al. 2014). They reported % improvements of 5.1–37.5% over 6 month follow-ups, which are not inconsistent with our results. However, comparison can be difficult because of variation between studies in terms of patient characteristics, diagnosis and severity.

Some studies have evaluated the effects CBT or group therapy on harmful events (Brown and Jager-Hyman 2014; Davidson et al. 2006, 2010; Slee et al. 2008; Stewart et al. 2009; Wood et al. 2001). Estimates of average instances of self-harm avoided (per person) varied from as little as 1.2 over 7 months to 3.4 over 3 months; for attempted suicides estimates ranged from 0.22 over 7 weeks to 0.53 over a year. In this light, our estimates of events avoided seem very high particularly for self-harm. Even though some of these studies exclude patients with certain categories of SMI, our results may reflect bias caused by the large amounts of attrition for this outcome.

It is conventional for traditional trial based evaluations to collect data for as many as 12 resource use categories—in contrast, data returns only reported HTT contacts, acute bed days and costs of therapies included in the IAPT program. Therefore, our analysis is likely to have underestimated costs in each arm. This resource use can include GP monitoring; contact with social care workers, mental health nurses and occupational therapists; accident and emergency services; drug and alcohol services; and referrals to secondary care services (e.g. appointment with a psychiatrist). The analysis tacitly assumes that any background treatments were similar between arms. This assumption would be invalid and overestimate incremental costs if IAPT were to displace other treatments. However, we have seen that our results are not broadly inconsistent with other research. Given the nature of routine data collection, the outcome measures used in this analysis do not capture all the benefits of psychological therapies such as recovery and overall wellbeing. Therefore, in that this analysis does not capture all relevant costs and benefits, it cannot be considered a comprehensive cost-effectiveness analysis of IAPT-SMI.

Nevertheless, we believe that this analysis makes a much-needed contribution to the literature because there have been no other economic evaluations of IAPT-SMI. Our results were not contradicted by estimates of incremental cost and WSAS from evaluations of individual disorders. There are no obvious policy implications from this evaluation. However, we believe greatest benefit will come from comparison with future studies of IAPT-SMI, which will help to validate our framework for analysing aggregate routine data.

In final conclusion, our analysis provides some evidence that IAPT-SMI may improve function and occurrence of harmful behaviours. In a similar way to conventional IAPT, there is evidence IAPT-SMI increases overall costs but saves non-psychological treatment costs over time. However, there is a need for more evaluations of IAPT-SMI using conventional and/or more robust study designs that evaluate a wider range of clinical outcomes.

Funding This work was supported by NHS England.

Compliance with Ethical Standards

Conflict of interest Author Darshan Zala declares that he has no conflict of interest. Author Paul McCrone declares that he has no conflict of interest. Author Muralikrishnan Radhakrishnan Kartha declares that he has no conflict of interest. Author Alison Brabban had been (or is currently) employed in some capacity by the English NHS. Author Alex Stirzaker had been (or is currently) employed in some capacity by the English NHS.

Ethical Approval was not required for this study because the article does not contain any studies with human participants performed by any of the authors. The data provided by NHS England that was used in the decision analytical model was collected routinely as part of service evaluation and was not individual level data.

Appendix 1

See Tables 2, 3 and 4.

Table 2 Model inputs into resource use and harmful event models for TAU arm

Model	Parameter	Psychosis or bipolar disorder				Personality disorder			
		BS	SLAM-EI	SLAM-PR	Lanc-EI	BEH	NEL	Som-PC	Som-SC
All	P(completers)	0.827	0.885	0.909	0.659	0.773	0.676	N/A	0.461
	P(drop-outs)	0.173	0.115	0.091	0.341	0.227	0.324	N/A	0.539
HTT	P(> 0 mean events | completers)	0.281	0.293	0.161	0.355	0.484	0.342	N/A	0.408
	Mean events| completers	2.280	0.020	0.020	1.190	1.000	1.710	N/A	0.890
	P(> 0 mean events | drop-outs)	0.286	0.333	0.304	0.333	0.317	0.437	N/A	0.478
	Mean events| drop-outs	2.770	0.130	0.080	1.610	1.380	1.930	N/A	0.950
Acute bed days	P(> 0 mean events | completers)	0.162	0.359	0.183	0.200	0.283	0.129	N/A	0.270
	Mean events| completers	5.320	0.050	0.010	5.400	0.980	2.750	N/A	3.320
	P(> 0 mean events | drop-outs)	0.229	0.583	0.348	0.193	0.139	0.206	N/A	0.304
	Mean events| drop-outs	2.850	0.460	0.250	3.550	1.550	2.320	N/A	0.720
Self-harm	P(> 0 mean events | completers)	N/A	N/A	N/A	N/A	0.942	0.837	N/A	0.442
	Mean events| completers	N/A	N/A	N/A	N/A	3.860	9.000	N/A	0.790
	P(> 0 mean events | drop-outs)	N/A	N/A	N/A	N/A	0.960	0.746	N/A	0.442
	Mean events| drop-outs	N/A	N/A	N/A	N/A	4.310	5.270	N/A	0.690
Suicide	P(> 0 mean events | completers)	N/A	N/A	N/A	N/A	0.895	0.464	N/A	0.195
	Mean events| completers	N/A	N/A	N/A	N/A	0.610	0.250	N/A	0.160
	P(> 0 mean events | drop-outs)	N/A	N/A	N/A	N/A	0.861	0.413	N/A	0.183
	Mean events| drop-outs	N/A	N/A	N/A	N/A	0.560	0.260	N/A	0.140
Violence	P(> 0 mean events | completers)	N/A	N/A	N/A	N/A	0.845	0.122	N/A	0.062
	Mean events| completers	N/A	N/A	N/A	N/A	0.550	1.970	N/A	0.170
	P(> 0 mean events | drop-outs)	N/A	N/A	N/A	N/A	0.822	0.119	N/A	0.087
	Mean events| drop-outs	N/A	N/A	N/A	N/A	0.480	1.930	N/A	0.080

No data was provided for harmful events for any of the 4 sites related to psychosis and bipolar disorder for the 12 month before period (i.e. TAU arm). No data was provided in the data return from Som-PC for either resource use or harmful events (IAPT arm)

Table 3 Model inputs into resource use and harmful event models for IAPT arm

Model	Parameter	Psychosis or bipolar disorder				Personality disorder			
		BS	SLAM-EI	SLAM-PR	Lanc-EI	BEH	NEL	Som-PC	Som-SC
All	P(completers)	0.820	0.885	0.909	0.685	0.750	0.588	N/A	N/A
	P(drop-outs)	0.180	0.115	0.091	0.315	0.250	0.412	N/A	N/A
HTT	P(> 0 mean events | completers)	0.102	0.076	0.043	0.240	0.404	0.127	N/A	N/A
	Mean events| completers	1.960	0.040	0.010	0.870	0.600	1.510	N/A	N/A
	P(> 0 mean events | drop-outs)	0.200	0.417	0.087	0.087	0.303	0.182	N/A	N/A
	Mean events| drop-outs	2.270	0.200	0.090	0.750	0.690	1.660	N/A	N/A
Acute bed days	P(> 0 mean events | completers)	0.124	0.217	0.065	0.160	0.172	0.051	N/A	N/A
	Mean events| completers	2.850	0.050	0.020	7.520	1.010	1.750	N/A	N/A
	P(> 0 mean events | drop-outs)	0.100	0.500	0.130	0.130	0.121	0.100	N/A	N/A
	Mean events| drop-outs	2.130	0.610	0.340	6.000	0.220	3.420	N/A	N/A
Self-harm	P(> 0 mean events | completers)	N/A	N/A	N/A	N/A	0.773	0.419	N/A	N/A
	Mean events| completers	N/A	N/A	N/A	N/A	0.500	2.090	N/A	N/A
	P(> 0 mean events | drop-outs)	N/A	N/A	N/A	N/A	0.833	0.500	N/A	N/A
	Mean events| drop-outs	N/A	N/A	N/A	N/A	0.130	4.210	N/A	N/A
Suicide	P(> 0 mean events | completers)	N/A	N/A	N/A	N/A	0.600	0.097	N/A	N/A
	Mean events| completers	N/A	N/A	N/A	N/A	0.060	0.170	N/A	N/A
	P(> 0 mean events | drop-outs)	N/A	N/A	N/A	N/A	0.462	0.125	N/A	N/A
	Mean events| drop-outs	N/A	N/A	N/A	N/A	0.050	0.500	N/A	N/A
Violence	P(> 0 mean events | completers)	N/A	N/A	N/A	N/A	0.398	0.097	N/A	N/A
	Mean events| completers	N/A	N/A	N/A	N/A	0.160	1.550	N/A	N/A
	P(> 0 mean events | drop-outs)	N/A	N/A	N/A	N/A	0.250	0.000	N/A	N/A
	Mean events| drop-outs	N/A	N/A	N/A	N/A	0.020	N/A	N/A	N/A

No data was provided for harmful events for any of the 4 sites related to psychosis and bipolar disorder for the 12 month after period (i.e. IAPT arm). No data was provided in the data returns from Som-PC and Som-SC for either resource use or harmful events (IAPT arm)

Table 4 Model inputs into WSAS model

Probability parameters	Psychosis or bipolar disorder				Personality disorder			
	BS	SLAM-EI	SLAM-PR	Lanc-EI	BEH	NEL	Som-PC	Som-SC
P (Worse 20% or less | completers)	0.000	0.000	0.000	0.000	0.000	0.057	0.001	0.000
P (Worse 21–40% | completers)	0.006	0.000	0.005	0.000	0.039	0.016	0.003	0.000
P (Worse 41–60% | completers)	0.000	0.051	0.015	0.000	0.039	0.049	0.010	0.000
P (Worse 61–80% | completers)	0.043	0.017	0.080	0.031	0.020	0.061	0.048	0.000
P (Worse 81–100% | completers)	0.230	0.136	0.206	0.125	0.353	0.154	0.199	0.426
P (no change | completers)	0.087	0.068	0.085	0.438	0.039	0.073	0.044	0.318
P (Improve 20% or less | completers)	0.354	0.407	0.307	0.234	0.333	0.146	0.326	0.256
P (Improve 21–40% | completers)	0.193	0.254	0.196	0.078	0.118	0.154	0.241	0.000
P (Improve 41–60% | completers)	0.081	0.034	0.085	0.047	0.059	0.114	0.097	0.000
P (Improve 61–80% | completers)	0.006	0.034	0.020	0.031	0.000	0.093	0.030	0.000
P (Improve 81–100% | completers)	0.000	0.000	0.000	0.016	0.000	0.081	0.002	0.000
P (Worse 20% or less | drop-outs)	0.000	N/A	N/A	N/A	0.000	0.101	0.000	0.000
P (Worse 21–40% | drop-outs)	0.000	N/A	N/A	N/A	0.100	0.030	0.011	0.000
P (Worse 41–60% | drop-outs)	0.037	N/A	N/A	N/A	0.000	0.051	0.011	0.000
P (Worse 61–80% | drop-outs)	0.074	N/A	N/A	N/A	0.200	0.051	0.057	0.000
P (Worse 81–100% | drop-outs)	0.296	N/A	N/A	N/A	0.300	0.111	0.307	0.309
P (no change | drop-outs)	0.037	N/A	N/A	N/A	0.100	0.051	0.102	0.564
P (Improve 20% or less | drop-outs)	0.296	N/A	N/A	N/A	0.100	0.192	0.307	0.127
P (Improve 21–40% | drop-outs)	0.185	N/A	N/A	N/A	0.100	0.192	0.170	0.000
P (Improve 41–60% | drop-outs)	0.074	N/A	N/A	N/A	0.100	0.061	0.034	0.000
P (Improve 61–80% | drop-outs)	0.000	N/A	N/A	N/A	0.000	0.081	0.000	0.000
P (Improve 81–100% | drop-outs)	0.000	N/A	N/A	N/A	0.000	0.081	0.000	0.000

Data returns for SLAM-EI, SLAM-PR and Lanc-EI did not contain data on WSAS change for patients classified as "drop-out

References

Barton, G. R., Hodgekins, J., Mugford, M., Jones, P. B., Croudace, T., & Fowler, D. (2009) Cognitive behaviour therapy for improving social recovery in psychosis: Cost-effectiveness analysis. *Schizophrenia Research, 112*(1–3), 158–163.

Briggs, A. H., Ades, A. E., & Price, M. J. (2003). Probabilistic sensitivity analysis for decision trees with multiple branches: Use of the Dirichlet distribution in a Bayesian framework. *Medical Decision Making, 23*(4), 341–350.

Brown, G. K., & Jager-Hyman, S. (2014). Evidence-based psychotherapies for suicide prevention: Future directions. *American Journal of Preventive Medicine, 47*(3 Suppl 2), S186–S194.

Charlwood, P., Mason, A., Goldacre, M., Cleary, R., & Wilkinson, E. (Eds.). (1999). Health outcome indicators: Severe mental illness. Report of a working group to the Department of Health. Oxford: National Centre for Health Outcomes Development. Retrieved from http://nchod.uhce.ox.ac.uk/mentalillness.pdf.

Curtis, L. (2015). PSSRU: unit costs of health and social care. Retrieved from http://www.pssru.ac.uk/project-pages/unit-costs/2015/.

Davidson, K., Norrie, J., Tyrer, P., Gumley, A., Tata, P., Murray, H., & Palmer, S. (2006). The effectiveness of cognitive behavior therapy for borderline personality disorder: Results from the borderline personality disorder study of cognitive therapy (BOSCOT) trial. *Journal of Personality Disorders, 20*(5), 450–465.

Davidson, K. M., Tyrer, P., Norrie, J., Palmer, S. J., & Tyrer, H. (2010). Cognitive therapy v. usual treatment for borderline personality disorder: Prospective 6-year follow-up. *The British Journal of Psychiatry, 197*(6), 456–462.

Davies, S. C. (2013). Chief Medical Officer's summary. In: N. Metha (ed.), Annual Report of the Chief Medical Officer 2013, Public Mental Health Priorities: Investing in the Evidence [online]. London: Department of Health, pp. 11–19. Retrieved from https://www.gov.uk/government/publications/chiefmedical-officer-cmo-annual-report-public-mental-health.

Durham, R. C., Chambers, J. A., Power, K. G., Sharp, D. M., Macdonald, R. R., Major, K. A., Dow, M. G., & Gumley, A. I. (2005). Long-term outcome of cognitive behaviour therapy clinical trials in central Scotland. *Health Technology Assessment, 9*(42), 1–174.

Durham, R. C., Guthrie, M., Morton, R. V., Reid, D. A., Treliving, L., Fowler, D., & Macdonald, R. R. (2003). Tayside–Fife clinical trial of cognitive-behaviour therapy for medication-resistant psychotic symptoms: results to three-month follow-up. *The British Journal of Psychiatry, 182*, 303–311.

Gumley, A. I., O'Grady, M., McNay, L., Reilly, J., Power, K. G., & Norrie, J. (2003). Early intervention for relapse in schizophrenia: Results of a 12-month randomised controlled trial of cognitive behavioural therapy. *Psychological Medicine, 33*, 419–431.

Jolley, S., Garety, P., Peters, E., Fornells-Ambrojo, M., Onwumere, J., Harris, V., Brabban, A., & Johns, L. (2015). Opportunities and challenges in Improving Access to Psychological Therapies for people with Severe Mental Illness (IAPT-SMI): evaluating the first operational year of the South London and Maudsley (SLaM) demonstration site for psychosis. *Behaviour Research and Therapy, 64*, 24–30.

Kohn, R., Saxena, S., Levav, I., & Saraceno, B. (2004). The treatment gap in mental health care. *Bulletin of the World Health Organization, 82*, 858–866.

Kuipers, E., Fowler, D., Garety, P., Chisholm, D., Freeman, D., Dunn, G., Bebbington, P., & Hadley, C. (1998). London-East Anglia randomised controlled trial of cognitive-behavioural therapy for psychosis. III: Follow-up and economic evaluation at 18 months. *The British Journal of Psychiatry, 173*, 61–68.

McCrone, P., Dhanasiri, S., Patel, A., Knapp, M., & Lawton-Smith, S. (2008). *Paying the price: The cost of mental health care in England to 2026*. London: King's Fund.

McPin, F. (2015). A service user evaluation of IAPT for people with a severe mental illness. Retrieved from http://mcpin.org/a-service-user-evaluation-of-iapt-for-people-with-a-severe-mental-illness/.

Mukuria, C., Brazier, J., Barkham, M., Connell, J., Hardy, G., Hutten, R., Saxon, D., Dent-Brown, K., & Parry, G. (2013). Cost-effectiveness of an improving access to psychological therapies service. *The British Journal of Psychiatry, 202*(3), 220–227.

Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The work and social adjustment scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry, 180*, 461–464.

Murphy, D., Hodgman, G., Carson, C., Spencer-Harper, L., Hinton, M., Wessely, S., & Busuttil, W. (2015). Mental health and functional impairment outcomes following a 6-week intensive treatment programme for UK military veterans with post-traumatic stress disorder (PTSD): A naturalistic study to explore dropout and health outcomes at follow-up. *BMJ Open. 5*(3), e007051.

National Centre for Social Research (2009). Adult psychiatric morbidity in England, 2007: Results of a household survey. Retrieved from http://www.esds.ac.uk/doc/6379/mrdoc/pdf/6379research_report.pdf.

NHS England. Adult Improving Access to Psychological Therapies programme. Retrieved from https://www.england.nhs.uk/mentalhealth/adults/iapt/.

NHS reference costs (2015). Retrieved from https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015.

Palmer, S., Davidson, K., Tyrer, P., Gumley, A., Tata, P., Norrie, J., Murray, H., & Seivewright, H. (2006). The cost-effectiveness of cognitive behavior therapy for borderline personality disorder: Results from the BOSCOT trial. *Journal of Personality Disorders, 20*(5), 466–481.

Patel, A., Knapp, M., Romeo, R., Reeder, C., Matthiasson, P., Everitt, B., & Wykes, T. (2010).Cognitive remediation therapy in schizophrenia: Cost-effectiveness analysis. *Schizophrenia Research*. 120(1–3), 217–224.

Petrou, S., & Gray, A. (2011). Economic evaluation using decision analytical modelling: Design, conduct, analysis and reporting. *BMJ, 11*(342), d1766.

Ridsdale, L., Hurley, M., King, M., McCrone, P., & Donaldson, N. (2012). The effect of counselling, graded exercise and usual care for people with chronic fatigue in primary care: A randomized trial. *Psychological Medicine, 42*(10), 2217–2224.

Slee, N., Garnefski, N., van der Leeden, R., Arensman, E., & Spinhoven, P. (2008). Cognitive-behavioural intervention for self-harm: Randomised controlled trial. *The British Journal of Psychiatry, 192*(3), 202–211.

Startup, M., Jackson, M. C., Evans, K. E., & Bendix, S. (2005). North Wales randomized controlled trial of cognitive behaviour therapy for acute schizophrenia spectrum disorders: Two-year follow-up and economic evaluation. *Psychological Medicine, 35*(9), 1307–1316.

Stewart, C. D., Quinn, A., Plever, S., & Emmerson, B. (2009). Comparing cognitive behavior therapy, problem solving therapy, and treatment as usual in a high risk population. *Suicide and Life-Threatening Behavior, 39*, 538–547.

Talbot, L. S., Maguen, S., Metzler, T. J., Schmitz, M., McCaslin, S. E., Richards, A., & Neylan, T. C. (2014). Cognitive behavioral therapy for insomnia in posttraumatic stress disorder: A randomized controlled trial. *Sleep, 37*(2), 327–341.

Wood, A., Trainor, G., Rothwell, J., Moore, A., & Harrington, R. (2001). Randomized trial of group therapy for repeated deliberate self-harm in adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(11), 1246–1253.

Zu, S., Xiang, Y. T., Liu, J., Zhang, L., Wang, G., Ma, X., Kilbourne, A. M., Ungvari, G. S., Chiu, H. F., Lai, K. Y., Wong, S. Y., Yu, D. S., & Li, Z. J. (2014). A comparison of cognitive-behavioral therapy, antidepressants, their combination and standard treatment for Chinese patients with moderate-severe major depressive disorders. *Journal of Affective Disorders, 152–154*, 262–267.