Behaviour Research and Therapy 79 (2016) 15–22

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Different people respond differently to therapy: A demonstration using patient profiling and risk stratification

Jaime Delgadillo [a, *], Omar Moreea [b], Wolfgang Lutz [c]

[a] Leeds Community Healthcare NHS Trust, and Department of Health Sciences, University of York, UK
[b] Centre for Clinical Practice, National Institute for Health and Care Excellence, UK
[c] Department of Psychology, University of Trier, Germany

ARTICLE INFO

Article history:
Received 16 July 2015
Received in revised form
24 December 2015
Accepted 18 February 2016
Available online 23 February 2016

Keywords:
IAPT
Patient profiling
Risk stratification
Depression
Anxiety

ABSTRACT

Background: This study aimed to identify patient characteristics associated with poor outcomes in psychological therapy, and to develop a patient profiling method.
Method: Clinical assessment data for 1347 outpatients was analysed. Final treatment outcome was based on reliable and clinically significant improvement (RCSI) in depression (PHQ-9) and anxiety (GAD-7) measures. Thirteen patient characteristics were explored as potential outcome predictors using logistic regression in a cross-validation design.
Results: Disability, employment status, age, functional impairment, baseline depression and outcome expectancy predicted post-treatment RCSI. Regression coefficients for these factors were used to derive a weighting scheme called Leeds Risk Index (LRI), used to assign risk scores to individual cases. After stratifying cases into three levels of LRI scores, we found significant differences in RCSI and treatment completion rates. Furthermore, LRI scores were significantly correlated with the proportion of treatment sessions classified as 'not on track'.
Conclusions: The LRI tool can identify cases at risk of poor progress to inform personalized treatment recommendations for low and high intensity psychological interventions.

1. Introduction

Although psychological interventions for mental health problems can be helpful for many people, not all patients have the same response to treatment. For example, researchers have observed that some patients (approximately between 15% and 45%) do not experience clinically significant improvement following treatment (Hansen, Lambert, & Forman, 2002) and up to 10% of cases actually deteriorate (Lambert & Ogles, 2004). Thus, it is important to find ways to identify and manage cases at risk of poor outcomes. This concern is at the heart of *patient focused research*, which seeks to develop decision rules and methods to enhance outcomes for individuals (Lutz, 2002). Two notable approaches within this line of research include outcome tracking and patient profiling.

Outcome tracking involves gathering relevant psychometric measures throughout treatment and using these to compare an individual patient's progress against normative data from a clinical population. Typically, data from cohorts of patients with the same intake scores on psychometric measures are aggregated to derive expected treatment response (ETR) norms (e.g. see Finch, Lambert, & Schaalje, 2001; Lambert et al., 2002; Lueger et al., 2001; Lutz, Martinovich, & Howard, 1999). Patients whose current scores denote a level of impairment which is outside of the ETR norms are classified as 'not on track' (NOT). Outcome 'feedback' involves alerting clinicians about cases that are identified as NOT, which can prompt a review of therapy and the application of clinical decision rules to prevent poor outcomes (Harmon et al., 2007; Whipple et al., 2003). A number of reviews and meta-analyses concur on the usefulness of outcome tracking and feedback as a means of improving outcomes for individual patients (Carlier et al., 2012; Castonguay, Barkham, Lutz, & McAleavey, 2013; Knaup, Koesters, Schoefer, Becker, & Puschner, 2009; Lambert et al., 2003; Shimokawa, Lambert, & Smart, 2010).

Patient profiling, on the other hand, involves predicting outcomes for individual patients based on their unique characteristics, presentation and life context. Patient profiling is founded on the observation that even patients with the same diagnosis can vary widely in other demographic and clinical characteristics (Garfield,

* Corresponding author. Mental Health and Addiction Research Group, Area 4, ARRC Building, Department of Health Sciences, University of York, Heslington YO10 5DD, UK.
E-mail address: jaime.delgadillo@nhs.net (J. Delgadillo).

http://dx.doi.org/10.1016/j.brat.2016.02.003
0005-7967/

1996; Kiesler, 1966). But just how important is variability in patient-factors when it comes to treatment outcomes? In a review on this subject, Garfield (1994) noted that baseline severity of psychopathology, pre-treatment expectancies and response during the early stages of therapy seemed to be plausible clinical outcome predictors. Since then, numerous other investigations have been published, examining the predictive utility of variables such as co-morbidity of mental disorders (Hoyer et al., 2014; Karlsson et al., 2008; Licht-Strunk et al., 2009; van Beljouw, Verhaak, Cuijpers, van Marwijk, & Penninx, 2010), personality disorders (Goddard, Wingrove, & Moran, 2015; Meyer, Pilkonis, Proietti, Heape, & Egan, 2001; Reich, 2003), baseline functioning and impairment (Frank et al., 2011), 'chronicity' or problem duration (Clark et al., 2009; Hamilton & Dobson, 2002; Karlsson et al., 2008; Richards & Borglin, 2011), family history of mental health problems (Dowrick et al., 2011a; Licht-Strunk et al., 2009), prior treatment episodes (Dobson et al., 2008; Lorenzo-Luaces, DeRubeis, & Webb, 2014; Lutz, Leon, Martinovich, Lyons, & Stiles, 2007), socioeconomic status (Self, Oates, Pinnock-Hamilton, & Leach, 2005), and pre-treatment expectancies (Constantino, Arnkoff, Glass, Ametrano, & Smith, 2011; Dowrick et al., 2011b; Grilo et al., 1998; Lutz et al., 2007; Swift & Callahan, 2011). In spite of the burgeoning research on patient-factors, it is still unclear how specific factors are weighted (e.g. strength of association) with respect to other possibly informative characteristics. Therefore, studies with large sets of variables and weighting schemes are necessary to advance this literature.

Practical applications of the above research findings are much less common. Bridging between the outcome tracking and patient profiling approaches, Lutz and collaborators (1999, 2001, and 2005) applied multilevel modelling in large clinical datasets to estimate individual patients' expected trajectory of improvement across sessions as a function of their pre-treatment clinical characteristics. More recently, DeRubeis et al. (2014) generated a personalized advantage index (PAI), which uses patient characteristics to ascertain which of two available treatments may be more advantageous to individual patients (DeRubeis et al., 2014; Huibers et al., 2015). Although the usefulness of the PAI method is yet to be tested prospectively, this important work presages the future possibility of individualized treatment recommendations. Still, further evidence is needed to determine which and how many pre-treatment variables are necessary to accurately predict treatment outcomes (Lutz et al., 2005; Lutz et al., 2006; Rubenstein et al., 2007).

With this backdrop of emerging evidence, the present study was based on three objectives. (1) To determine the prognostic accuracy of several patient characteristics gathered as part of intake assessments in a primary care mental health setting. (2) To construct a patient profiling tool that could be used in routine practice. (3) To examine the clinical utility of the patient profiling tool.

2. Methods

2.1. Setting and participants

This study used anonymous clinical records for 1347 outpatients who accessed psychological treatment in a primary care mental health service in the North of England which was aligned to the national *Improving Access to Psychological Therapies* (IAPT) programme. IAPT services offer a range of evidence-based interventions for depression and anxiety organised in a stepped care model (Clark et al., 2009) in accordance with clinical guidelines (National Institute for Health and Care Excellence, 2011). In this model low intensity treatments (Step 2 in the treatment pathway) are offered as a starting point for patients with mild-to-moderate conditions; these involve teaching and supporting patients to apply self-help strategies based on cognitive behavioural therapy (CBT) principles. These interventions typically last between one and eight sessions, they rely on didactic materials, and are supported by qualified mental health practitioners. High intensity psychological interventions are considered the next step up in the model (Step 3); they are often lengthier (e.g. up to 20 sessions) and are offered to those who have not derived benefit from low intensity care or those with more severe clinical presentations. Step 3 interventions included CBT, interpersonal psychotherapy, EMDR, and counselling for depression.

The mean age in the sample was 37.9 (SD = 14.2); 65.6% were females; 89.8% were of a White British background; and 49.4% were unemployed (43.3% working, 7.3% full-time students). Primary diagnoses were established through semi-structured interviews supplemented by screening tools for depression and anxiety disorders (IAPT National Programme Team, 2011). The most common primary problems were major depression (35.1%), mixed anxiety and depressive disorder (36.4%), generalized anxiety disorder (12.2%), panic disorder (5.7%), obsessive compulsive disorder (3.4%), social phobia (2.3%), post-traumatic stress disorder (1.3%), with other problems accounting for less than 4% of recorded diagnoses. Approximately 67.6% of patients in this sample were only treated at Step 2, and 32.4% accessed Step 3 interventions (67.8% of whom had CBT).

2.2. Measures and data sources

Two symptom questionnaires were taken as primary outcome measures, which are consistent with the service's target population. The PHQ-9 is a nine-item screening tool for major depressive disorder (Kroenke, Spitzer, & Williams, 2001). Each item is rated on a 0 to 3 scale, yielding a total depression severity score between 0 and 27. A diagnostic cut-off $\geq$ 10 has been recommended for the detection of major depression, with adequate sensitivity (88%) and specificity (88%). GAD-7 is a seven-item questionnaire developed to screen for anxiety disorders (Spitzer, Kroenke, Williams, & Lowe, 2006). Each item is also rated on a 0 to 3 ordinal scale, rendering a total severity score between 0 and 21. A cut-off score $\geq$ 8 in this questionnaire is recommended to identify the likely presence of an anxiety disorder with adequate sensitivity (77%) and specificity (82%) (Kroenke, Spitzer, Williams, Monahan, & Löwe, 2007). Both questionnaires were self-completed by patients on a session-to-session basis to monitor progress during therapy, and the last observed measures (for completers and dropouts) were used to assess final treatment outcomes using *intention-to-treat* principles. In this dataset, dropout was defined as a unilateral decision by the patient to stop attending treatment, and these cases were identified by a specific variable contained in clinical records.

Potential outcome predictors were grouped into demographic and clinical factors. All variables were derived from information gathered as part of 45-min semi-structured assessment interviews conducted when patients initially accessed the service. Demographics included: age; gender; ethnicity; employment status (employed vs. unemployed); socio-economic status. Following Paddison et al. (2012), socio-economic status was derived by matching each patient's home postcode to the Index of Multiple Deprivation (Department for Communities and Local Government, 2011) and generating a 5-level ordinal variable where higher levels denoted greater deprivation. Clinical factors characterised the history and profile of the patient's condition; these included family history of psychiatric problems; chronicity of mental health problems (in years and months); number of prior treatment episodes; chronic physical illness; self-reported disability; and outcome expectancy. Outcome expectancy was measured using a single question based on Lutz et al. (2007), rated on a Likert scale (range

0 = low expectancy, to 10 = high expectancy). Clinical factors also included baseline symptom severity (PHQ-9, GAD-7) and psychosocial functioning measured using the Work and Social Adjustment Scale (WSAS) (Mundt, Mark, Shear, & Griest, 2002).

2.3. Statistical analyses

The analyses were performed in 3 stages which aimed to (1) identify outcome predictors, (2) to develop a patient profiling and risk index, and (3) to evaluate the predictive utility of the risk index.

The first stage involved multivariate logistic regression modelling based on a backward elimination method. The outcome in these models was a binary variable which indicated whether or not patients met criteria for reliable and clinically significant improvement (RCSI) at the end of treatment, based on the method outlined by Jacobson and Truax (1991). The outcome variable was coded 0 = RCSI and 1 = no RCSI, so that the models would predict the risk of a poor treatment outcome. Separate models for depression (PHQ-9) and anxiety (GAD-7) were developed applying the diagnostic cut-offs and reliable change indices described by Delgadillo et al. (2014). Spurious results can be obtained in large samples when examining several variables (Altman, Gore, Gardner, & Pocock, 1983); therefore we applied a series of methods to enhance the reliability of the analysis. The dataset was randomly split into an estimation sample (N = 651) and a validation sample (N = 696). The analysis was initially performed in the estimation sample entering all potential predictors, and replicated 1000 times using bootstrap resampling (Mick & Ratain, 1994). *Goodness-of-fit* in these analyses was assessed using the Hosmer–Lemeshow test (Lemeshow & Hosmer, 1982). Residual plots were also examined to diagnose the adequacy of regression models. Using backward elimination, the final estimation model only retained predictors which were significant at a level of $p < .05$. The next step involved testing the predictive validity of the parameters of the estimation model in the validation dataset. Any predictors that were not supported in the validation sample were removed until a final model was obtained.

The next stage aimed to construct a risk stratification method based on the patient characteristics that were found to predict poor outcomes. Any continuous outcome predictors identified in the prior stage were transformed into categorical variables split by quartiles. This was necessary to be able to classify or 'group' patients with different levels of risk. Depression (PHQ-9) and anxiety (GAD-7) logistic regression models (as described above) were run in the full sample entering the categorical outcome predictors corresponding to each of the respective outcome measures. Weights for each variable were determined based on the coefficients from the final regression models. Following the risk stratification method outlined by Fan et al. (2002), the beta coefficient for each variable was multiplied by four and then rounded to the nearest integer to derive a simplified risk weighting scheme. The resulting weighting scheme combined all variables from depression and anxiety regression models, which we refer to as the Leeds Risk Index (LRI). Each case in the full sample was then assigned an LRI score (range = 0–21) matched to their baseline characteristics, where higher LRI scores indicated an accumulation of risk factors. The rationale for developing an index was to be able to rate the potential risk of poor progress for individual patients using a simple and intuitive scale.

In the final stage of analysis, we aimed to empirically test the clinical utility of the LRI. To do this, we formulated a series of questions which would enable us to confirm or disconfirm our assumptions about patient profiling. (1) Are patients with higher LRI scores less likely to attain RCSI after therapy? (2) Are LRI scores associated with dropout from therapy? (3) Do LRI scores still predict final treatment outcome after controlling for baseline severity, early response to treatment and length of treatment? (4) Are higher LRI scores associated with the probability of being classed as 'not on track' (NOT) during the course of therapy?

To answer question 1, we stratified all cases into different risk categories according to their LRI scores, and used bar charts to visually compare the % of cases with RCSI across categories. Initially we split the sample into quartiles; however the two middle quartiles had comparable RCSI rates so we merged them to arrive at a three-group split that optimally differentiated cases. We applied chi-square analyses to formally compare RCSI rates across categories. We also used bar charts to compare RCSI rates attained at step 2 and step 3 of the stepped care treatment pathway. To answer question 2, we used bar charts and chi-square analyses to compare the number of cases that dropped out of treatment across risk categories. To answer question 3, we estimated change scores in PHQ-9 and GAD-7 measures between the intake assessment and the third therapy session (indicative of early response to treatment). We then entered LRI scores into logistic regression models predicting RCSI post-treatment, additionally controlling for baseline severity, early change scores and total number of therapy sessions attended.

To answer the fourth question, PHQ-9 and GAD-7 measures gathered on a session-to-session basis for each case were analysed using hierarchical linear modelling to develop expected treatment response (ETR) curves following the method outlined by Finch et al. (2001). Mean predicted values derived from these regression models and 80% confidence intervals were calculated for each treatment session, for every cluster of patients with a shared intake score on the relevant outcome measure. Observed outcome scores at each therapy session were then classified as 'on track' (OT) if they were smaller than the upper-boundary of the confidence intervals (representing the 90 th percentile marker), or classified as 'not on track' (NOT) if larger than this boundary. This enabled us to code every single session in the dataset as either OT or NOT; the exception to this was for a small proportion of cases (N = 21; 1.6% of 1347 cases) that had missing intake scores for one of the two outcome measures and for whom we were unable to generate ETR curves. We then estimated a single index for each case that represented the proportion of treatment sessions classed as NOT (% NOT) in each of the two outcome measures. In the final steps of this analysis, we used Spearman's (non-parametric) correlations to assess whether LRI scores were associated with %NOT variables for PHQ-9 and GAD-7. A sensitivity analysis was carried out applying partial correlations to assess whether associations between LRI and %NOT variables were statistically significant when controlling for the total number of treatment sessions. This was considered important because differences in therapy length obviously affect the weighting of the %NOT variable.

3. Results

3.1. Investigation of outcome predictors

The baseline levels of depression and anxiety symptoms in the study sample were PHQ = 15.91 (SD = 5.66) and GAD-7 = 14.24 (SD = 4.30). A total of 1168 cases met criteria for clinically significant depression symptoms (PHQ-9 ≥ 10), of whom 543 (46.5%) had RCSI at the end of treatment. Of the 1275 cases with clinically significant anxiety symptoms (GAD ≥ 8), a total of 595 (46.7%) had RCSI.

Using a backward elimination logistic regression method applied in the estimation sample, we arrived at a parsimonious model which explained approximately 15% of variance in depression (PHQ-9) outcomes (Nagelkerke $R^2 = .15$). This model correctly classified 63.4% of cases based on pre-treatment variables. Results

indicated that younger age (B = −.37, SE = .13, β = .69, $p < .01$), unemployment (B = .69, SE = .18, β = 1.99, $p < .001$), having a self-reported disability (B = .79, SE = .22, β = 2.21, $p < .001$), and higher baseline impairment on WSAS (B = .06, SE = .01, β = 1.06, $p < .001$) significantly predicted a poor treatment outcome. The results of bootstrap resampling did not differ from the predictors obtained without resampling. When the parameters obtained in the estimation model were tested in the validation dataset, all of the same variables were statistically significant predictors of a poor depression outcome. The final models from the estimation and validation samples are presented in Table 1.

Using the same method, the estimation model for GAD-7 explained approximately 13% of variance (Nagelkerke R^2 = .13) and indicated that higher baseline depression (PHQ-9) severity (B = .07, SE = .02, β = 1.07, $p < .001$), higher baseline impairment on WSAS (B = .03, SE = .01, β = 1.03, $p = .03$), having a family history of mental health problems (B = .42, SE = .17, β = 1.52, $p = .02$), greater number of prior treatment episodes (B = .11, SE = .05, β = 1.11, $p = .03$), and lower outcome expectancy (B = −.10, SE = .05, β = .90, $p = .03$) all predicted a poor anxiety outcome. This model correctly classified 62.1% of cases. Results of bootstrap resampling were no different to those without resampling. However, the final model obtained in the validation sample only provided support for a narrower set of anxiety outcome predictors, these were: higher baseline depression (PHQ-9) severity (B = .06, SE = .02, β = 1.06, $p = .001$) and lower outcome expectancy (B = −.12, SE = .05, β = .89, $p = .01$). These models are presented in Table 2.

3.2. Development of a risk weighting scheme

As in the prior stage of iterative regression analysis, several models were constructed, but this time using predictors that were transformed to categorical variables. We aimed to obtain final models that maximized predictive power (Nagelkerke R^2 statistic) and optimized *goodness-of-fit* (Hosmer–Lemeshow test). We found that the optimal models required dichotomizing the age variable into "teenager vs. adult" categories (using an age cut-off < 20) instead of quintile or quartile groups, as well as splitting the expectancy variable into 3 groups. The resulting optimal models based on categorical variables are presented in Table 3. This table also shows the LRI patient profiling method, which assigns a 'score' (weight) to specific characteristics. For example, a teenage patient (weight = 3) with severe baseline depression (PHQ-9 = 21 to 27 = weight = 6), moderate functional impairment (WSAS = 21 to 30 = weight = 3) and low expectations of improvement (Expectancy = 0 to 5 = weight = 3) would have a total LRI score of 15 (sum of weights = 3 + 6 +3 + 3). In this way, many possible combinations of patient characteristics (e.g. different profiles) can be stratified according to low (0–4), moderate (5–9), or high (10–21) risk.

3.3. Evaluating the clinical utility of the Leeds Risk Index

Fig. 1 shows that RCSI rates were significantly lower in the groups with moderate and high LRI scores. This trend was confirmed for both Step 2 (PHQ-9: $x^2 = 20.80$, DF = 2, $p < .001$; GAD-7: $x^2 = 40.04$, DF = 2, $p < .001$) and Step 3 (PHQ-9: $x^2 = 16.00$, DF = 2, $p < .001$; GAD-7: $x^2 = 14.50$, DF = 2, $p < .001$) interventions. We noticed an exception where Step 2 cases with low and moderate LRI scores had comparable RCSI rates (42%), but cases with high LRI scores had considerably poorer outcomes (22% RCSI). Overall, Step 3 interventions had higher RCSI rates across all LRI categories. The relative advantage of Step 3 interventions (9%–23% difference) was more pronounced for depression cases with low LRI scores, and anxiety cases with moderate and high LRI scores. Fig. 2 shows treatment completion rates across LRI categories, also comparing Step 2 and Step 3 interventions. Differences between LRI categories were statistically significant ($x^2 = 23.58$, DF = 2, $p < .001$) in the full sample; this was primarily because cases with high LRI scores had considerably lower treatment completion rates (55.2%) compared to those with low (71.2%) and moderate scores (69.3%). Treatment completion rates were generally higher for Step 3, though the

Table 1
Cross-validation of depression (PHQ-9) outcome prediction models.

	First estimation sample model				Final estimation sample model				Final validation sample model			
	Nagelkerke R^2 = .177 Hosmer–Lemeshow χ^2 = 4.060, p = .852				Nagelkerke R^2 = .150 Hosmer–Lemeshow χ^2 = 1.589, p = .991				Nagelkerke R^2 = .071 Hosmer–Lemeshow χ^2 = 8.032, p = .430			
Variables	B	SE	β	p	B	SE	β	p	B	SE	β	p
Constant	−1.011	.671	.364	.132	−.840	.400	.432	.036	−.167	.379	.846	.659
Age (quintiles)	−.390	.135	.677	**.004**	−.374	.126	.688	**.003**	−.287	.122	.750	**.018**
Unemployed (ref = employed)	.638	.188	1.893	**.001**	.689	.184	1.992	**<.001**	.492	.173	1.636	**.004**
Disabled (ref = no disability)	.675	.227	1.965	**.003**	.793	.215	2.209	**<.001**	.553	.202	1.739	**.006**
Baseline WSAS	.048	.013	1.049	**<.001**	.059	.011	1.061	**<.001**	.031	.011	1.031	**.004**
Gender (ref = male)	−.190	.193	.827	.324								
Ethnicity: white (ref)				.941								
Ethnicity: mixed	.156	.513	1.168	.762								
Ethnicity: South Asian	−.068	.506	.935	.894								
Ethnicity: Black	.534	.545	1.705	.327								
Ethnicity: Chinese	.501	1.298	1.650	.700								
Ethnicity: Other	.252	1.056	1.287	.811								
SES (quintiles)	.081	.072	1.084	.259								
LTC (ref = no LTC)	.010	.209	1.010	.961								
Family history (ref = no history)	.128	.189	1.137	.497								
Chronicity (in months)	.000	.001	1.000	.842								
Prior treatment episodes	.070	.048	1.073	.139								
Baseline PHQ-9	.049	.026	1.050	.065								
Baseline GAD-7	−.012	.025	.988	.619								
Expectancy	−.072	.052	.930	.164								

Dependent variable = poor PHQ-9 outcome defined as not meeting criteria for reliable and clinically significant improvement after therapy; ref = reference category; SES = Socio-economic status rank derived from multiple deprivation index (quintile groups); B = unstandardized regression coefficient; SE = standard error of B; β = odds ratio; LTC = long term condition (co-morbid physical illnesses); statistically significant probability (p) values highlighted in **bold** text.

Table 2
Cross-validation of anxiety (GAD-7) outcome prediction models.

	First estimation sample model				Final estimation sample model				Final validation sample model			
	Nagelkerke $R^2 = .152$ Hosmer–Lemeshow $\chi^2 = 9.810$, $p = .279$				Nagelkerke $R^2 = .127$ Hosmer–Lemeshow $\chi^2 = 6.622$, $p = .578$				Nagelkerke $R^2 = .096$ Hosmer–Lemeshow $\chi^2 = 4.941$, $p = .764$			
Variables	B	SE	β	*p*	B	SE	β	*p*	B	SE	β	*p*
Constant	−.459	.614	.632	.454	−.702	.513	.496	.171	−.492	.454	.611	.278
Baseline PHQ-9	.078	.021	1.081	**<.001**	.067	.019	1.070	**<.001**	.061	.018	1.063	**.001**
Expectancy	−.097	.049	.908	**.047**	−.102	.047	.903	**.032**	−.117	.046	.889	**.011**
Baseline WSAS	.030	.012	1.030	**.015**	.026	.012	1.027	**.028**	.021	.012	1.021	.078
Family history (ref = no history)	.453	.179	1.573	**.011**	.420	.174	1.521	**.016**	.211	.171	1.235	.217
Prior treatment episodes	.097	.048	1.102	**.041**	.105	.047	1.111	**.025**	.045	.051	1.046	.377
Age (quintiles)	−.276	.127	.759	**.030**								
Unemployed (ref = employed)	.259	.178	1.295	.147								
Disabled (ref = no disability)	.246	.220	1.279	.263								
Gender (ref = male)	.048	.185	1.049	.796								
Ethnicity: white (ref)				.631								
Ethnicity: mixed	.933	.528	2.542	.077								
Ethnicity: South Asian	.041	.483	1.042	.932								
Ethnicity: Black	.305	.502	1.357	.543								
Ethnicity: Chinese	−.052	1.060	.949	.961								
Ethnicity: Other	.153	.805	1.166	.849								
SES (quintiles)	.021	.067	1.021	.756								
LTC (ref = no LTC)	.202	.198	1.224	.307								
Chronicity (in months)	.000	.001	1.000	.769								
Baseline GAD-7	−.048	.028	.954	.085								

Dependent variable = poor GAD-7 outcome defined as not meeting criteria for reliable and clinically significant improvement after therapy; ref = reference category; SES = Socio-economic status rank derived from multiple deprivation index (quintile groups); B = unstandardized regression coefficient; SE = standard error of B; β = odds ratio; LTC = long term condition (co-morbid physical illnesses); statistically significant probability (*p*) values highlighted in **bold** text..

difference was more pronounced for cases with high LRI scores (17% difference).

LRI scores remained significantly associated with final depression (B = .16, SE = .03, β = 1.18, $p < .001$) and anxiety (B = .15, SE = .02, β = 1.17, $p < .001$) outcomes, after controlling for baseline severity, early response and total therapy sessions using logistic regression models. These models accounted for 32% (9% unique to LRI) and 29% (4% unique to LRI) of variance in final PHQ-9 and GAD-7 outcomes respectively. These models correctly classified 71.9% of depression outcomes and 71.3% of anxiety outcomes. Finally, LRI scores were significantly correlated with the proportion of treatment sessions classified as 'not on track' (%NOT) using expected treatment response curves for PHQ-9 ($r = .10$, $p < .001$) and GAD-7 ($r = .20$, $p < .001$). These associations remained statistically significant when controlling for length of treatment in partial correlations between LRI and PHQ-9 ($r = .11$, $p < .001$), and LRI and GAD-7 ($r = .25$, $p < .001$). This indicates that patients with higher LRI scores are more likely to be classed as being not on track during therapy, and therefore have a different and less favourable trajectory of symptom changes.

4. Discussion

4.1. Main findings

This study supports the notion that different people respond

Table 3
Leeds Risk Index (LRI) patient profiling and risk stratification method.

	PHQ-9 outcome prediction model				GAD-7 outcome prediction model				Leeds risk Index
	Nagelkerke $R^2 = .088$ Hosmer–Lemeshow $\chi^2 = 2.278$, $p = .943$				Nagelkerke $R^2 = .092$ Hosmer–Lemeshow $\chi^2 = 6.277$, $p = .508$				Weighted risk scores for different
Variables	B	SE	β	*p*	B	SE	β	*p*	patient characteristics
Constant	−1.064	.217	.345	<.001	−.689	.167	.502	<.001	
Disabled (ref = no disability)	.583	.143	1.792	<.001					2
Unemployed (ref = employed)	.500	.126	1.648	<.001					2
Teenager (ref = adult with age ≥ 20)	.652	.284	1.920	.022					3
Baseline WSAS = 0–10 (ref)				<.001					0
Baseline WSAS = 11–20	.580	.221	1.785	.009					2
Baseline WSAS = 21–30	.869	.221	2.385	<.001					3
Baseline WSAS = 31–40	1.316	.262	3.729	<.001					5
Baseline PHQ-9 = 0–9 (ref)								<.001	0
Baseline PHQ-9 = 10–15					.385	.186	1.469	.39	2
Baseline PHQ-9 = 16–20					.679	.187	1.972	<.001	3
Baseline PHQ-9 = 21–27					1.425	.202	4.158	<.001	6
Expectancy = 8–10 (ref)								<.001	0
Expectancy = 6–7					.108	.132	1.114	.413	0
Expectancy = 0–5					.633	.158	1.883	<.001	3

B = unstandardized regression coefficient; SE = standard error of B; β = odds ratio; the reference categories for binary variables (e.g. no disability, employed, adults) should be assigned a score of 0; the LRI method helps to rate the cumulative risk of poor outcomes on a continuous scale ranging between 0 and 21; risk level can also be stratified according to low (0–4), moderate (5–9), or high (10–21) LRI scores.

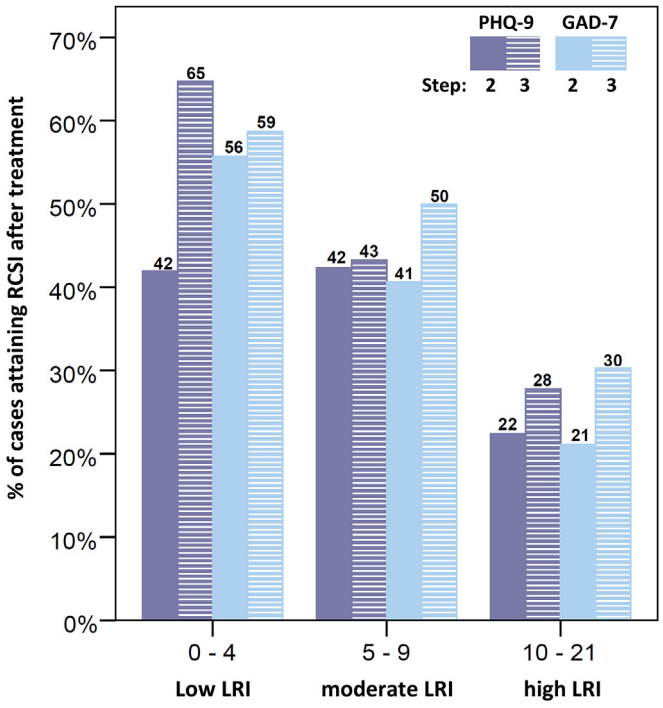

Fig. 1. Leeds Risk Index (LRI) as a predictor of reliable and clinically significant improvement (RCSI).

differently to psychological therapy, based on their individual characteristics and circumstances. Our results indicate that disability, unemployment, younger age (<20) and functional impairment are associated with the persistence of depressed mood after therapy. Furthermore, co-morbid depression symptoms and low expectations about the potential benefit of therapy appear to be associated with the persistence of anxiety symptoms after

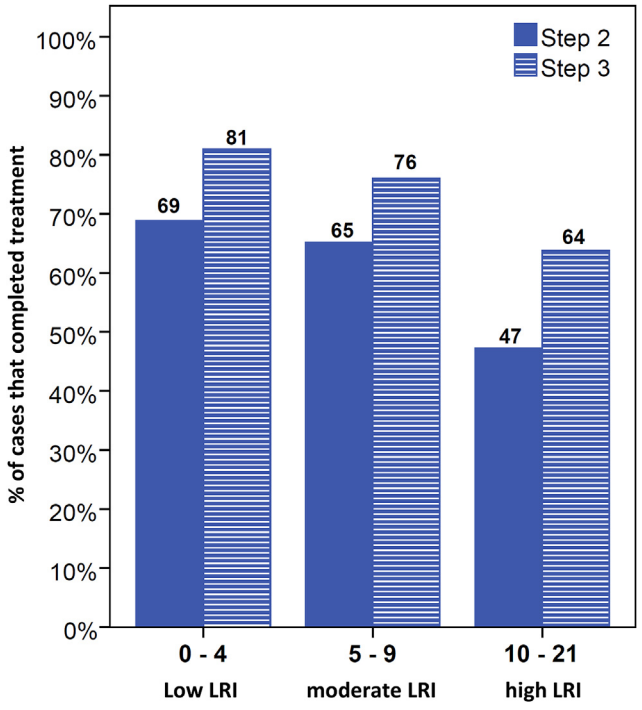

Fig. 2. Leeds Risk Index (LRI) as a predictor of treatment completion.

treatment. Symptoms measured by PHQ-9 and GAD-7 are highly correlated and likely to measure a common aspect of psychopathology (Böhnke, Lutz, & Delgadillo, 2014). We therefore developed a patient profiling method that combines all of the above risk factors for these types of symptoms. The resulting LRI method can enable clinicians to rate the risk of poor treatment outcomes for individual cases, and to stratify this into low, moderate or high risk categories. A particular advantage of employing a risk index is that it uses a relatively simple and intuitive scale that can classify numerous possible combinations of patient characteristics and associated regression weights.

We propose that LRI information can improve our ability to predict clinical outcomes. The estimation of probability of a particular outcome is generally more accurate when it is informed by base-rates, rather than relying on intuition or personal experience (Tversky & Kahneman, 1974). Applying this logic, an educated guess before starting treatment would be that any particular patient has a 46% probability of attaining RCSI, since this is the base-rate in this sample. But predictions can be improved upon the base-rate if we have statistically valid information about individual cases (Kahneman, 2011). Using the LRI helps us to calibrate our prognosis for cases whose probability of improvement is markedly lower than the 'average patient': cases with high LRI scores (≥10) had considerably low RCSI rates (average of 25%) across Step 2 (low intensity) and Step 3 (high intensity) interventions. These cases also tend to have persistently high symptom scores (classed as 'not on track') during treatment, and are at increased risk of dropout and poor final outcomes. Furthermore, LRI scores continued to significantly predict up to 9% of variance in outcomes after controlling for baseline severity, early response and therapy length.

4.2. Strengths and limitations

The outcome prediction analysis in this study was enhanced by the application of well established cross-validation and resampling methods (see Sauerbrei, 1999). The large dataset enabled the replication of findings in a statistically independent validation sample to prevent reaching conclusions based on spurious associations. A series of empirical tests enabled us to verify our assumptions about cumulative risk factors, also applying growth curve modelling methods to assess the utility of the LRI within the outcome tracking and feedback paradigm. Of particular note is the simultaneous analysis of more than a dozen patient characteristics. Compared to most of the studies cited in the introduction, which assessed the predictive accuracy of small sets of variables, this dataset enabled us to identify key outcome predictors whilst controlling for various potentially confounding variables.

As is common in large naturalistic datasets, we encountered missing data points for self-reported outcome measures. For example, a small number (N = 21) of cases only had baseline scores for one of the outcome measures. This meant that we could not produce growth curves or pre-post treatment outcomes for these cases on at least one measure, but they were nevertheless included in analysis using the available outcome and assessment data.

In spite of the inclusion of over a thousand cases with complete assessment data, the Step 3 sample size was insufficient to carry out cross-validated outcome prediction analyses for specific high intensity treatments. The ongoing accumulation of this type of data should enable more fine-grained analyses to understand which predictors may interact with which treatments or combinations of interventions. The LRI could also be combined with additional variables for this purpose; for example including personality disorder measures (see Goddard et al., 2015) to investigate the possibility of developing personalized advantage indices for different treatment modalities (DeRubeis et al., 2014; Huibers et al., 2015).

This dataset did not contain information about the therapists that treated each case, so we were unable to assess variability in clinical outcomes that may be attributable to therapists. Nevertheless, we note that a recent study conducted in a similar setting (Firth, Barkham, Kellett, & Saxon, 2015) found that similar patient-variables (symptom severity, functioning, age, employment status) differentially moderate clinical outcomes even after controlling for variability between therapists (e.g. *therapist effects*). We consider that this convergence of findings by independent research groups enhances the reliability of the evidence base. Nevertheless, we recognise that the LRI method requires further replication in other datasets, particularly since it was developed using data from a single service.

4.3. Considerations for clinical practice

The LRI method could be feasibly integrated into a spread-sheet or a web-application to help therapists to easily classify patients with different profiles using a 'traffic light' system (low, moderate, high risk). This risk stratification system could support personalized clinical decision making in stepped care services. Depression cases with low LRI and anxiety disorder cases with moderate LRI have an advantage if treated at Step 3. However, they are likely to complete treatment, so from a cost-efficiency perspective it is appropriate to offer Step 2 interventions initially, expecting that the smaller proportion of cases who require Step 3 are likely to attend and to gain further improvements. Depression cases with moderate LRI and anxiety disorder cases with low LRI attain similar outcomes in Step 2 or Step 3, and are likely to complete treatment. This clinical equipoise justifies offering conventional stepped care. On the contrary, depression and anxiety disorder cases with high LRI have a modest advantage (6%–9% RCSI difference) at Step 3, but the probability of dropout in these cases is particularly high at Step 2 (>50%). On this basis, we recommend *patient-intensity-matching*, where these cases are referred directly to Step 3.

Patient-intensity-matching could potentially enhance completion and clinical outcomes for high risk cases. Ideally, the cost-effectiveness of this strategy should be tested prospectively using experimental or quasi-experimental designs using historical controls. In addition to this, we recommend that close monitoring of early response to treatment could considerably improve the accuracy of identifying cases at risk of poor outcomes. It may be especially important to monitor early response to therapy in cases with high LRI scores, since they are more likely to be classed as 'not on track' and may therefore benefit most from outcome feedback strategies (Shimokawa et al., 2010).

Declarations of interest

None.

Acknowledgements

This study was supported and funded by Leeds Community Healthcare NHS Trust, United Kingdom. Ethical approval for this study was provided by an NHS Research Ethics Committee, as part of a programme of work on clinical outcome prediction (Reference: 10/H1306/68). The funding organisation had no influence in the design, analysis or decision to publish the findings of this study.

References

Altman, D. G., Gore, S. M., Gardner, M. J., & Pocock, S. J. (1983). Statistical guidelines for contributors to medical journals. *British Medical Journal, 286*(6376), 1489–1493.

van Beljouw, I. M., Verhaak, P. F., Cuijpers, P., van Marwijk, H. W., et al. (2010). The course of untreated anxiety and depression, and determinants of poor one-year outcome: a one-year cohort study. *BMC Psychiatry, 10*(1), 86.

Böhnke, J. R., Lutz, W., & Delgadillo, J. (2014). Negative affectivity as a trans-diagnostic factor in patients with common mental disorders. *Journal of Affective Disorders, 166*, 270–278.

Carlier, I. V. E., Meuldijk, D., Van Vliet, I. M., Van Fenema, E., Van der Wee, N. J. A., & Zitman, F. G. (2012). Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *Journal of Evaluation in Clinical Practice, 18*, 104–110.

Castonguay, L. G., Barkham, M., Lutz, W., & McAleavey, A. A. (2013). Practice-oriented research: approaches and applications. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change* (6th ed.). New York: Wiley & Sons.

Clark, D. M., Layard, R., Smithies, R., Richards, D. A., Suckling, R., & Wright, B. (2009). Improving access to psychological therapy: initial evaluation of two UK demonstration sites. *Behaviour Research and Therapy, 47*(11), 910–920.

Constantino, M. J., Arnkoff, D. B., Glass, C. R., Ametrano, R. M., & Smith, J. Z. (2011). Expectations. *Journal of Clinical Psychology: In Session, 67*, 184–192.

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early changes, attrition and dose-response in low intensity psychological interventions. *British Journal of Clinical Psychology, 53*(1), 114–130.

Department for Communities and Local Government. (2011). *The English indices of deprivation 2010*. London: DCLG. Retrieved from http://www.reigate-banstead.gov.uk/Images/English%20Indices%20of%20Deprivation%202010%20Note_tcm9-44047.pdf.

DeRubeis, R. J., Cohen, Z. D., Forand, N. R., Fournier, J. C., Gelfand, L. A., & Lorenzo-Luaces, L. (2014). The personalized advantage index: translating research on prediction into individualized treatment recommendations. *A demonstration*. http://dx.doi.org/10.1371/journal.pone.0083875.

Dobson, K. S., Hollon, S. D., Dimidjian, S., Schmaling, K. B., Kohlenberg, R. J., Gallop, R., et al. (2008). Randomized trial of behavioural activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology, 76*(3), 468–477.

Dowrick, C., Flach, C., Leese, M., Chatwin, J., Morriss, R., Peveler, R., et al. (2011b). Estimating probability of sustained recovery from mild to moderate depression in primary care: evidence from the THREAD study. *Psychological Medicine, 41*(1), 141–150.

Dowrick, C., Shiels, C., Page, H., Ayuso-Mateos, J. L., Casey, P., Dalgard, O. S., et al. (2011a). Predicting long-term recovery from depression in community settings in Western Europe: evidence from ODIN. *Social Psychiatry and Psychiatric Epidemiology, 46*(2), 119–126.

Fan, V. S., Au, D., Heagerty, P., Deyo, R. A., McDonell, M. B., & Fihn, S. D. (2002). Validation of case-mix measures derived from self-reports of diagnoses and health. *Journal of Clinical Epidemiology, 55*(4), 371–380.

Finch, A. E., Lambert, M. J., & Schaalje, B. G. (2001). Psychotherapy quality control: the statistical generation of expected recovery curves for integration into an early warning system. *Clinical Psychology and Psychotherapy, 8*(4), 231–242.

Firth, N., Barkham, M., Kellett, S., & Saxon, D. (2015). Therapist effects and moderators of effectiveness and efficiency in psychological wellbeing practitioners: a multilevel modelling analysis. *Behaviour Research and Therapy, 69*, 54–62.

Frank, E., Cassano, G. B., Rucci, P., Thompson, W. K., Kraemer, H. C., Fagiolini, A., et al. (2011). Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. *Psychological Medicine, 41*, 151–162.

Garfield, S. L. (1994). Research on client variables in psychotherapy. In A. E. Bergin, & S. L. Garfield (Eds.), *Handbook of psychotherapy and behavior change* (4th ed.). New York: Wiley & Sons.

Garfield, S. L. (1996). Some problems associated with "validated" forms of psychotherapy. *Clinical Psychology: Science and Practice, 3*, 218–229.

Goddard, E., Wingrove, J., & Moran, P. (2015). The impact of comorbid personality difficulties on response to IAPT treatment for depression and anxiety. *Behaviour Research and Therapy, 73*, 1–7.

Grilo, C. M., Money, R., Barlow, D. H., Goddard, A. W., Gorman, J. M., Hofmann, S. G., et al. (1998). Pretreatment patient factors predicting attrition from a multicenter randomized controlled treatment study for panic disorder. *Comprehensive Psychiatry, 39*(6), 323–332.

Hamilton, K. E., & Dobson, K. S. (2002). Cognitive therapy of depression: pretreatment patient predictors of outcome. *Clinical Psychology Review, 22*(6), 875–893.

Hansen, N. B., Lambert, M. J., & Forman, E. M. (2002). The psychotherapy dose–response effect and its implications for treatment delivery services. *Clinical Psychology: Science and Practice, 9*, 329–343.

Harmon, S. C., Lambert, M. J., Smart, D. M., Hawkins, E., Nielsen, S. L., Slade, K., et al. (2007). Enhancing outcome for potential treatment failures: therapist– client feedback and clinical support tools. *Psychotherapy Research, 17*(4), 379–392.

Hoyer, J., Wiltink, J., Hiller, W., Miller, R., Salzer, S., Sarnowsky, S., et al. (2014). Baseline patient characteristics predicting outcome and attrition in cognitive therapy for social phobia: results from a large multicentre trial. *Clinical Psychology & Psychotherapy*. http://dx.doi.org/10.1002/cpp.1936. advance online publication.

Huibers, M. J. H., Cohen, Z. D., Lemmens, L. H. J. M., Arntz, A., Peeters, F. P. M. L., Cuijpers, P., et al. (2015). Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized

advantage index approach. *PLoS One, 10*(11), e0140771. http://dx.doi.org/10.1371/journal.pone.0140771.

IAPT National Programme Team. (2011). *The IAPT Data Handbook: Guidance on recording and monitoring outcomes to support local evidence-based practice*. Version 2.0. London: National IAPT Programme Team. Retrieved from http://www.iapt.nhs.uk/silo/files/iapt-datahandbook.pdf.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux (Chapter 16).

Karlsson, L., Kiviruusu, O., Miettunen, J., Heilä, H., Holi, M., Ruuttu, T., et al. (2008). One-year course and predictors of outcome of adolescent depression: a case-control study in Finland. *Journal of Clinical Psychiatry, 69*(5), 844–853.

Kiesler, D. J. (1966). Some myths of psychotherapy research and the search for a paradigm. *Psychological Bulletin, 65*, 110–136.

Knaup, C., Koesters, M., Schoefer, D., Becker, T., & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *British Journal of Pyschiatry, 195*, 15–22.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606–613.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine, 146*(5), 317–325.

Lambert, M. J., & Ogles, B. M. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change* (5th ed., pp. 139–193). New York: Wiley.

Lambert, M. J., Whipple, J. L., Bishop, M. J., Vermeersch, D. A., Gray, G. V., & Finch, A. E. (2002). Comparison of empirically-derived and rationally-derived methods for identifying patients at risk for treatment failure. *Clinical Psychology and Psychotherapy, 9*(3), 149–164.

Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10*(3), 288–301.

Lemeshow, S., & Hosmer, D. W. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. *American Journal of Epidemiology, 115*(1), 92–106.

Licht-Strunk, E., Van Marwijk, H. W. J., Hoekstra, T., Twisk, J. W. R., De Haan, M., & Beekman, A. T. F. (2009). Outcome of depression in later life in primary care: longitudinal cohort study with three years' follow-up. *British Medical Journal, 338*. http://dx.doi.org/10.1136/bmj.a3079.

Lorenzo-Luaces, L., DeRubeis, R. J., & Webb, C. A. (2014). Client characteristics as moderators of the relation between the therapeutic alliance and outcome in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 82*(2), 368–373.

Lueger, R. J., Howard, K. I., Martinovich, Z., Lutz, W., Anderson, E. E., & Grissom, G. (2001). Assessing treatment progress of individual patients using expected treatment response models. *Journal of Consulting and Clinical Psychology, 69*(2), 150–158.

Lutz, W. (2002). Patient-focused psychotherapy research and individual treatment progress as scientific groundwork for an empirically based clinical practice. *Psychotherapy Research, 12*(3), 251–272.

Lutz, W., Barkham, M., Stiles, W., Noble, R., Iveson, S., Leach, C., et al. (2005). Predicting change for individual psychotherapy clients on the basis of their nearest neighbors. *Journal of Consulting and Clinical Psychology, 73*(5), 904–913.

Lutz, W., Lambert, M. J., Harmon, S. C., Tschitsaz, A., Schürch, E., & Stulz, N. (2006). The probability of treatment success, failure and duration - what can be learned from empirical data to support decision making in clinical practice? *Clinical Psychology and Psychotherapy, 13*(4), 223–232.

Lutz, W., Leon, S. C., Martinovich, Z., Lyons, J. S., & Stiles, W. B. (2007). Therapist effects in outpatient psychotherapy: a three level growth curve approach. *Journal of Counseling Psychology, 54*(1), 32–39.

Lutz, W., Martinovich, Z., & Howard, K. I. (1999). Patient profiling: an application of random coefficient regression models to depicting the response of a patient to outpatient psychotherapy. *Journal of Consulting and Clinical Psychology, 67*(4), 571–577.

Meyer, B., Pilkonis, P. A., Proietti, J. M., Heape, C. L., & Egan, M. (2001). Attachment styles and personality disorders as predictors of symptom course. *Journal of Personality Disorders, 15*(5), 371–389.

Mick, R., & Ratain, M. J. (1994). Bootstrap validation of pharmacodynamic models defined via stepwise linear regression. *Clinical Pharmacology and Therapeutics, 56*(2), 217–222.

Mundt, J. C., Mark, I. M., Shear, M. K., & Griest, J. M. (2002). The work and social adjustment scale: a simple measure of impairment in functioning. *British Journal of Psychiatry, 180*(5), 461–464.

National Institute for Health and Care Excellence. (2011). *Common mental health disorders: Identification and pathways to care* [CG123]. London: National Institute for Health and Care Excellence. Retrieved from http://www.nice.org.uk/guidance/CG123.

Paddison, C., Elliott, M., Parker, R., Staetsky, L., Lyratzopoulos, G., Campbell, J. L., et al. (2012). Should measures of patient experience in primary care be adjusted for case mix? Evidence from the english general practice patient survey. *BMJ Quality and Safety, 21*(8), 634–640.

Reich, J. (2003). The effect of axis II disorders on the outcome of treatment of anxiety and unipolar depressive disorders: a review. *Journal of Personality Disorders, 17*(5), 387–405.

Richards, D. A., & Borglin, G. (2011). Implementation of psychological therapies for anxiety and depression in routine practice: two year prospective cohort study. *Journal of Affective Disorders, 133*(1), 51–60.

Rubenstein, L. V., Rayburn, N. R., Keeler, E. B., Ford, D. E., Rost, K. M., & Sherbourne, C. D. (2007). Predicting outcomes of primary care patients with major depression: development of a depression prognosis index. *Psychiatric Services, 58*(8), 1049–1056.

Sauerbrei, W. (1999). The use of resampling methods to simplify regression models in medical statistics. *Applied Statistics, 48*(3), 313–329.

Self, R., Oates, P., Pinnock-Hamilton, T., & Leach, C. (2005). The relationship between social deprivation and unilateral termination (attrition) from psychotherapy at various stages of the health care pathway. *Psychology and Psychotherapy: Theory, Research and Practice, 78*(1), 95–111.

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*(3), 298–311.

Spitzer, R., Kroenke, K., Williams, J. B. W., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine, 166*(10), 1092–1097.

Swift, J. K., & Callahan, J. L. (2011). Decreasing treatment dropout by addressing expectations for treatment length. *Psychotherapy Research, 21*(2), 193–200.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124–1131.

Whipple, J. L., Lambert, M. J., Vermeersch, D. A., Smart, D. W., Nielsen, S. L., & Hawkins, E. J. (2003). Improving the effects of psychotherapy: the use of early identification of treatment failure and problem-solving strategies in routine practice. *Journal of Counseling Psychology, 50*(1), 59–68.