Psychological Medicine (2012), **42**, 967–980. © Cambridge University Press 2011
doi:10.1017/S0033291711001905

ORIGINAL ARTICLE

Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms

R. Uher[1]*, R. H. Perlis[2], N. Henigsberg[3], A. Zobel[4], M. Rietschel[5], O. Mors[6], J. Hauser[7], M. Z. Dernovsek[8], D. Souery[9], M. Bajs[3], W. Maier[4], K. J. Aitchison[1], A. Farmer[1] and P. McGuffin[1]

[1] *MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, UK*
[2] *Center for Experimental Drugs and Diagnostics, Department of Psychiatry and Center for Human Genetic Research, Massachusetts General Hospital, Boston, USA*
[3] *Croatian Institute for Brain Research, Medical School, University of Zagreb, Croatia*
[4] *Department of Psychiatry, University of Bonn, Germany*
[5] *Central Institute of Mental Health, Division of Genetic Epidemiology in Psychiatry, Mannheim, Germany*
[6] *Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Denmark*
[7] *Laboratory of Psychiatric Genetics, Department of Psychiatry, Poznan University of Medical Sciences, Poland*
[8] *University Psychiatric Clinic, Ljubljana, Slovenia*
[9] *Laboratoire de Psychologie Médicale, Université Libre de Bruxelles and Psy Pluriel – Centre Européen de Psychologie Médicale, Belgium*

Background. Symptom dimensions have not yet been comprehensively tested as predictors of the substantial heterogeneity in outcomes of antidepressant treatment in major depressive disorder.

Method. We tested nine symptom dimensions derived from a previously published factor analysis of depression rating scales as predictors of outcome in 811 adults with moderate to severe depression treated with flexibly dosed escitalopram or nortriptyline in Genome-based Therapeutic Drugs for Depression (GENDEP). The effects of symptom dimensions were tested in mixed-effect regression models that controlled for overall initial depression severity, age, sex and recruitment centre. Significant results were tested for replicability in 3637 adult out-patients with non-psychotic major depression treated with citalopram in level I of Sequenced Treatment Alternatives to Relieve Depression (STAR*D).

Results. The interest-activity symptom dimension (reflecting low interest, reduced activity, indecisiveness and lack of enjoyment) at baseline strongly predicted poor treatment outcome in GENDEP, irrespective of overall depression severity, antidepressant type and outcome measure used. The prediction of poor treatment outcome by the interest-activity dimension was robustly replicated in STAR*D, independent of a comprehensive list of baseline covariates.

Conclusions. Loss of interest, diminished activity and inability to make decisions predict poor outcome of antidepressant treatment even after adjustment for overall depression severity and other clinical covariates. The prominence of such symptoms may require additional treatment strategies and should be accounted for in future investigations of antidepressant response.

Received 7 July 2011; Revised 12 August 2011; Accepted 23 August 2011; First published online 20 September 2011

Key words: Antidepressant medication, depressive symptoms, dimensional classification, major depression, outcome prediction.

Introduction

More than 20 antidepressant drugs are available for treating depression, but outcomes of treatment are highly variable individually (Rush *et al*. 2006; Uher *et al*. 2010). Psychiatrists have made numerous attempts to define subtypes of depression that would be more homogeneous in response to treatment (Carney *et al*. 1965; Paykel *et al*. 1982; Fava *et al*. 1997; Parker *et al*. 1999). Although melancholic (Perry, 1996), atypical (Joyce *et al*. 2004) and anxious (Fava *et al*. 2008) depression predict treatment outcomes in some cohorts, inconsistent results have been reported for each (McGrath *et al*. 2000; Russell *et al*. 2001; Brown, 2007; McGrath *et al*. 2008; Thase, 2009; Nelson, 2010; Stewart *et al*. 2010; Uher *et al*. 2011). Therefore, more reliable predictors of individual differences in response to treatment are needed.

* Address for correspondence: Dr R. Uher, P080, SGDP, Institute of Psychiatry, 16 De Crespigny Park, London SE5 8AF, UK.
(Email: rudolf.uher@kcl.ac.uk)

Several authors have suggested that heterogeneity of depression may be better characterized by continuous dimensions than by categorical constructs (Flett *et al*. 1997; Prisciandaro & Roberts, 2009). Compared to categorical subtypes, dimensional classifications of depression are more reliable (Flett *et al*. 1997; Prisciandaro & Roberts, 2005; Parker *et al*. 2009) and their validity has been established in family (Korszun *et al*. 2004), epidemiological (Bjelland *et al*. 2009; Prisciandaro & Roberts, 2009) and biological investigations (Veen *et al*. 2011; Wardenaar *et al*. 2011). However, with few notable exceptions (Carney *et al*. 1965; Fava *et al*. 2008; Howland *et al*. 2008), dimensional symptom measures other than overall depression severity have not been investigated systematically as predictors of treatment outcome. The aim of this report was to explore the predictive validity of previously established symptom dimensions in a large sample of subjects with major depression (Uher *et al*. 2009*a*) and to establish the generalizability of significant predictors by replication in another large treatment sample (Trivedi *et al*. 2006). We hypothesized that symptom dimensions would predict treatment outcome and that the predictions would replicate across clinical populations.

Method

Patients

The discovery dataset was the Genome-based Therapeutic Drugs for Depression (GENDEP), a 12-week open-label part-randomized multi-centre study with two active pharmacological treatment arms (Uher *et al*. 2009*a*). It comprises 811 treatment-seeking adults diagnosed with ICD-10/DSM-IV unipolar major depression of at least moderate severity established in the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview (Wing *et al*. 1998), recruited in nine European centres. Personal or family history of bipolar disorder or schizophrenia and active substance dependence constituted exclusion criteria. The study was approved by ethics boards in all centres. All participants provided written consent after the procedures were explained. GENDEP is registered at EudraCT (No. 2004-001723-38, http://eudract.emea.europa.eu) and ISRCTN (No. 03693000, www.controlled-trials.com). A detailed description of the GENDEP sample is available elsewhere (Uher *et al*. 2009*a*) and in Supplementary Table S1 (available online).

The replication sample was the limited access dataset (version 2) distributed from the National Institutes of Health (NIH)-supported Sequenced Treatment Alternatives to Relieve Depression (STAR*D). The primary purpose of STAR*D was to determine which treatments work best if the first antidepressant treatment does not produce remission. The STAR*D sample comprises 4041 treatment-seeking adult out-patients with DSM-IV non-psychotic major depression, recruited in 31 centres in the USA. This study uses 3637 subjects with at least one measurement during citalopram treatment. The study was approved by institutional ethics review boards in participating centres. All participants provided written consent after the procedures and associated risks were explained. STAR*D is registered at ClinicalTrials.gov (NCT00021528). A detailed description of the STAR*D sample and design is available elsewhere (Rush *et al*. 2006; Trivedi *et al*. 2006) and in Supplementary Table S1.

Interventions

In GENDEP (Uher *et al*. 2009*a*), subjects were allocated to one of two antidepressants with different primary modes of action: escitalopram, a selective inhibitor of the serotonin transporter (SSRI), and nortriptyline, a tricyclic antidepressant inhibiting the noradrenaline transporter. Participants for whom the two antidepressants were at equipoise were randomly allocated to receive escitalopram or nortriptyline: 233 were randomized to escitalopram and 235 to nortriptyline. Patients with contra-indications for one of the drugs were allocated non-randomly to the other antidepressant: 225 to escitalopram and 118 to nortriptyline. Escitalopram was initiated at 10 mg daily and increased to a target dose of 15 mg daily within 2 weeks, with an optional increase to 20 mg daily, reaching a mean daily dose of 15.2 mg (S.D. = 7.3 mg) at study exit. Nortriptyline was initiated at 50 mg daily and titrated to a target dose of 100 mg daily within 2 weeks, with an optional increase to 150 mg daily, reaching a mean daily dose of 99.4 mg (S.D. = 37.6 mg) at study exit. Compliance was monitored by weekly self-reported pill count and plasma levels of antidepressants measured at week 8. Of the 811 participants, 628 (77%) completed 8 weeks and 527 (65%) completed 12 weeks on the allocated antidepressant. Individuals treated with escitalopram and nortriptyline improved to a similar degree (Uher *et al*. 2009*a*).

In STAR*D level 1, all subjects were treated with citalopram, a SSRI, for up to 14 weeks (Trivedi *et al*. 2006). The protocol-guided citalopram dose titration started with 20 mg daily, increased to 40 mg daily by week 4 and to 60 mg daily by week 6. The dose was individually adjusted according to the ratio of benefits to adverse effects in a measurement-based care framework (Trivedi *et al*. 2006). The mean dose of citalopram at study-level exit was 42 mg daily. The protocol recommended consultations with treating physicians at weeks 2, 4, 6, 9 and 12, and an optional

https://doi.org/10.1017/S0033291711001905 Published online by Cambridge University Press

final consultation at week 14. Per protocol, participants could exit this study level if they experienced intolerable adverse effects of citalopram, after 9 weeks of treatment with maximum tolerated dose or if they achieved remission, that is a score of $\leqslant 7$ on the Hamilton Rating Scale for Depression (HAMD-17; Hamilton, 1967). Of the 4041 subjects, 3637 (90.0%) had at least one post-baseline follow-up, 3054 (75.6%) were still in the study after 6 weeks, 2636 (65.2%) were treated for 9 weeks, and 2011 (49.8%) remained on citalopram for 12 weeks or longer.

Outcome measures

Given the nature of the predictors tested in this report and the fact that symptom dimensions may show differential correlation with various depression rating scales, we required *a priori* that prediction should be robust across clinician-rated and self-reported outcomes. Both studies used multiple depression-rating scales, including observer-rated and self-report instruments, to measure depression severity at study baseline, on multiple occasions during the treatment and at study/level exit.

In GENDEP, depression severity was measured weekly with three established scales: the clinician-rated Montgomery–Åsberg Depression Rating Scale (MADRS; Montgomery & Åsberg, 1979), the HAMD-17 (Hamilton, 1967) and the self-report Beck Depression Inventory (BDI; Beck *et al*. 1961). MADRS and HAMD-17 were administered by trained psychologists and psychiatrists with high inter-rater reliability (Uher *et al*. 2008). MADRS was the primary outcome measure. At least one valid post-baseline MADRS was available for 789 (97.3%), HAMD-17 for 791 (97.5%) and BDI for 781 (96.3%) subjects.

In STAR*D, depression severity was measured with HAMD-17, administered by independent research outcome assessors in telephone interviews at baseline and study/level exit. The clinician-rated (QIDS-C) and self-report (QIDS-SR) versions of the 16-item Quick Inventory of Depression Symptomatology (QIDS) were administered at each clinical visit (Rush *et al*. 2003). HAMD-17 was the planned primary outcome measure, but more complete data were obtained for QIDS-C and QIDS-SR, which were used as primary outcomes in most STAR*D reports (Trivedi *et al*. 2006; Fava *et al*. 2008). Valid post-baseline HAMD-17 was available for 2796 (69.2%), QIDS-C for 3630 (89.8%) and QIDS-SR for 3607 (89.3%) subjects.

Predictors

In GENDEP, continuous factor scores based on a previously published item-response categorical factor analysis of MADRS, BDI and HAMD-17 items were tested as potential predictors of outcome (Uher *et al*. 2008). Based on a parallel analysis, it was estimated that up to six factors were needed to describe the structure of depressive symptoms (Uher *et al*. 2008). The factor analysis identified three major factors (observed mood, cognitive and neurovegetative symptoms) that further split into six specific dimensions (mood, anxiety, pessimism, interest-activity, sleep and appetite; Fig. 1). Standardized dimension scores were obtained from a graded item-response theory model of symptoms with positive loading, so that higher scores on each dimension corresponded to more severe symptoms (Uher *et al*. 2008). Item response parameters for the interest-activity dimension are given in Supplementary Table S2.

In STAR*D, continuous scores matching the dimensions identified as significant predictors in GENDEP were constructed as sums of items with corresponding content on baseline HAMD-17, QIDS-SR, QIDS-C and the research outcome assessor-rated 30-item Inventory for Depression Symptomatology (IDS; Rush *et al*. 1996) (Supplementary Table S3).

Statistical analysis

The nine dimensional symptom scores (three major factors and six specific dimensions) were tested as predictors of continuous outcomes in GENDEP. Symptom dimensions that significantly predicted outcome in GENDEP were then tested for replicability in STAR*D. To establish that the results are robust to the choice of outcome measure and independent of overall depression severity, convergent results for three outcome scales were required and all analyses were controlled for the severity at baseline.

The effects of continuous predictors on response to antidepressants were tested using linear mixed-effect models fitted with maximum likelihood, as described previously (Uher *et al*. 2009*a*, *b*). Available data on depression severity at all post-baseline measurement occasions were included in the analyses. MADRS, HAMD-17 and BDI total scores were used as outcomes in GENDEP. HAMD-17, QIDS-SR and QIDS-C total scores were used as outcomes in STAR*D. The baseline total score on the outcome scale was always included as a covariate to account for general depression severity. Age, sex and linear and quadratic effects of time were also included as covariates in all analyses. Hierarchical random effects accounted for clustering of repeated measurements within individuals and clustering of individuals within recruiting clinics, ensuring that any results are independent of centre effects.

In GENDEP, each predictor was first tested in the whole sample, then within each treatment arm, and

https://doi.org/10.1017/S0033291711001905 Published online by Cambridge University Press

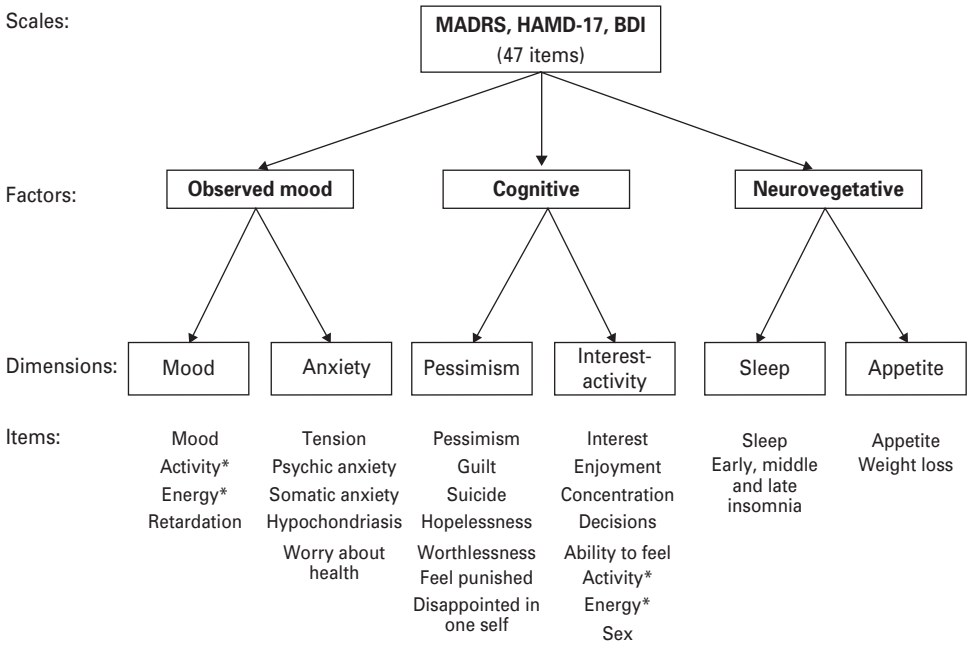

Fig. 1. Depression symptom structure. The figure reviews the results of categorical-factor analysis of items from three depression rating scales: the Montgomery–Åsberg Depression Rating Scale (MADRS), the 17-item Hamilton Rating Scale for Depression (HAMD-17) and the Beck Depression Inventory (BDI). Categorical item factor analysis identified three major factors: observed mood, cognitive symptoms and neurovegetative symptoms. These three factors further split into six dimensions: mood, anxiety, pessimism, interest-activity, sleep and appetite. * The items measuring activity and energy loaded on the observed mood factor in the three-factor solution, but in the six-factor solution, these items cross-loaded evenly between the mood and the interest-activity dimensions.

finally in interaction with the drug (escitalopram *versus* nortriptyline).

In the STAR*D study, significant predictors of outcome in the primary reports were included as covariates (Trivedi *et al*. 2006; Fava *et al*. 2008).

To correct for testing nine predictors in GENDEP, only predictors associated with outcome on the primary measure at $\alpha < 0.005$ and with congruent results on all three outcome measures were considered for replication in STAR*D. The significance threshold for replication was set at $0.05/n$, where n is the number of predictors tested for replication. As convergent findings on all outcome measures were required (rather than just a significant outcome on any measure), further correction for the number of outcome measures was not needed.

Clinical significance

Continuous outcomes and mixed-effect linear models were preferred for the primary analyses for reasons of statistical power, ability to control for bias and handle missing data (Deyi *et al*. 1998; Streiner, 2002; Mallinckrodt *et al*. 2010). However, the size of effects detected in these models may be difficult to interpret and apply to clinical practice. Therefore, we also computed the outcome of remission, using the widely accepted definition of a score of $\leqslant 7$ on HAMD-17 at the last clinical visit (Frank *et al*. 1991) and we present the direct relationship between this outcome and any significant predictor tested with a simple univariate logistic regression, without any covariates. A clinically meaningful measure of effect size is the number needed to treat (NNT), which can be more accurately referred to as the number needed to assess (NNA) for the purposes of prediction, and reflects the number of individuals who need to undergo an assessment for one additional treatment outcome to be accurately predicted. NNT/NNA can be computed for a relationship between continuous and categorical variables using the area under the receiver-operating characteristic (ROC) curve, the AUC (Kraemer & Kupfer, 2006). NNA has a straightforward meaning. For example, as the rate of remission in STAR*D is 41%, the best guess in the absence of a predictor would be that a given individual will not achieve remission. However, this guess will be wrong in four out of 10 individuals. A low NNA means that a relatively high proportion of individuals will be correctly reclassified and the prediction of outcome will be more accurate. For example, a predictor with an NNA of 5 will help to accurately predict remission in an

https://doi.org/10.1017/S0033291711001905 Published online by Cambridge University Press

additional two out of every 10 individuals, reducing the error of guessing by half. Although a simple threshold for 'significance' would be an over-simplification, an NNA smaller than 10 may recommend a low-burden test for clinical use (Kraemer & Kupfer, 2006).

Results

Three major symptom factors as predictors of outcome in GENDEP

The observed mood factor was significantly associated with older age at study entry (Spearman's $\rho = 0.11$, $p = 0.0014$) and later age of depression onset (Spearman's $\rho = 0.14$, $p = 0.0001$) but not with other baseline and treatment characteristics (sex, age, marital status, employment, episode duration, antidepressant treatment history, attrition or dose of either antidepressant; all $p > 0.05$). Higher observed mood scores at baseline predicted worse outcome of treatment on all three scales (Table 1), with the strongest effect on BDI. The effect was independent of drug and confirmed in sensitivity analyses incorporating additional covariates, including age of onset.

The cognitive symptom factor score was unrelated to other baseline characteristics (all $p > 0.05$) but was associated with higher exit dose of escitalopram (Spearman's $\rho = 0.22$, $p < 0.0001$) and higher exit dose of nortriptyline (Spearman's $\rho = 0.16$, $p = 0.0074$). Higher baseline cognitive symptom scores strongly predicted significantly worse outcome of treatment on MADRS and HAMD-17, but not on BDI (Table 1). Similar results were obtained in sensitivity analyses with additional covariates.

The neurovegetative symptom factor score was not a significant predictor of outcome (Table 1).

Six specific symptom dimensions as predictors of outcome in GENDEP

Four of the six specific symptom dimensions significantly predicted outcome on MADRS (Table 2). The interest and activity dimension was the strongest predictor and predicted outcome on each of the three depression rating scales at $p < 0.0001$, independently of overall baseline severity and of which antidepressant was used (Table 2). These effects were confirmed in sensitivity analyses restricted to randomly allocated individuals [e.g. for interest-activity and MADRS: $\beta = 0.21$, 95% confidence interval (CI) 0.13–0.27, $p = 7.0 \times 10^{-9}$]. Higher interest-activity scores were associated with later depression onset (Spearman's $\rho = 0.07$, $p = 0.0486$) and more previous depressive episodes (Spearman's $\rho = 0.08$, $p = 0.0211$) but no other baseline variables (all $p > 0.05$). The interest-activity dimension was also associated with a higher dose of escitalopram (Spearman's $\rho = 0.21$, $p = 0.0001$) and of nortriptyline (Spearman's $\rho = 0.18$, $p = 0.0033$). The prediction of outcome by the interest-activity dimension remained unchanged after controlling for potential confounders, including age of onset, number of depressive episodes and dose of antidepressants ($\beta = 0.18$, 95% CI 0.13–0.24, $p = 4.7 \times 10^{-10}$).

The only symptom dimension with evidence of differential prediction by drug was anxiety. Higher baseline scores on the anxiety dimension predicted worse outcome with nortriptyline but slightly better outcome with escitalopram (i.e. the effects were in the opposite directions in the two medication groups; interaction $p = 0.0233$; Table 2).

Specificity of prediction by the interest-activity symptom dimension in GENDEP

As the two symptom dimensions most predictive of outcome (interest-activity and mood, Table 2) shared cross-loaded items measuring activity and energy, we explored their relative contributions in an additional analysis with both interest-activity and mood dimensions entered as predictors of the primary outcome. This showed that the strong effect of the interest-activity dimension was independent of mood ($\beta = 0.18$, 95% CI 0.12–0.24, $p = 3.4 \times 10^{-10}$) and that the mood dimension did not carry additional predictive information independent of interest-activity ($p > 0.1$). This result confirmed the decision that only the interest-activity dimension should be followed up in the replication sample.

*Replication of the interest-activity dimension as a predictor of outcome in STAR*D*

The interest-activity symptom dimension fulfilled *a priori* criteria for pursuing replication (association at the corrected $p < 0.005$ in primary analysis and concordant results with all outcome measures). An equivalent score in STAR*D was constructed by summing HAMD-17, IDS and QIDS items corresponding to items forming the interest-activity score (Supplementary Table S3). Items with equivalent content and source (clinician *versus* self-report) were identified for all items, except that no equivalent to the self-reported work/activity on BDI was identified in STAR*D. The resulting scores were normally distributed (Supplementary Fig. S1).

A higher baseline interest-activity symptom score significantly predicted worse outcome of treatment with citalopram on all three outcome scales in STAR*D after correcting for overall baseline severity (all $p < 0.001$; Table 3, model A).

https://doi.org/10.1017/S0033291711001905 Published online by Cambridge University Press

Table 1. *Prediction of treatment outcome from the three baseline symptom dimensions in GENDEP*

	Overall		Escitalopram		Nortriptyline		Interaction with drug	
	($n=811$)		($n=457$)		($n=354$)		($n=811$)	
Predictor	β (95% CI)	p	β (95% CI)	p	β (95% CI)	p	β (95% CI)	p
(1) Outcome: MADRS								
Observed mood	0.11 (0.02 to 0.18)	**0.0134**	0.13 (0.02 to 0.23)	**0.0236**	0.08 (−0.05 to 0.19)	0.2338	0.00 (−0.09 to 0.07)	0.8468
Cognitive	0.14 (0.09 to 0.19)	**5.2×10^{-8}**	0.15 (0.08 to 0.21)	**3.3×10^{-5}**	0.13 (0.05 to 0.19)	**0.0013**	−0.01 (−0.10 to 0.05)	0.5394
Neurovegetative	0.00 (−0.05 to 0.05)	0.9730	0.00 (−0.07 to 0.07)	0.9541	0.00 (−0.07 to 0.08)	0.9728	0.00 (−0.08 to 0.08)	0.9293
(2) Outcome: HAMD-17								
Observed mood	0.15 (0.08 to 0.20)	**2.5×10^{-6}**	0.15 (0.06 to 0.22)	**0.0004**	0.14 (0.04 to 0.21)	**0.0042**	0.00 (−0.08 to 0.08)	0.9787
Cognitive	0.14 (0.08 to 0.18)	**2.2×10^{-8}**	0.14 (0.07 to 0.20)	**1.9×10^{-5}**	0.12 (0.04 to 0.18)	**0.0015**	−0.01 (−0.10 to 0.05)	0.5501
Neurovegetative	0.00 (−0.05 to 0.05)	0.9936	0.01 (−0.06 to 0.08)	0.8063	−0.01 (−0.09 to 0.06)	0.7328	−0.01 (−0.10 to 0.06)	0.6731
(3) Outcome: BDI								
Observed mood	0.12 (0.07 to 0.16)	**2.3×10^{-6}**	0.12 (0.05 to 0.18)	**0.0006**	0.13 (0.05 to 0.19)	**0.0006**	0.00 (−0.09 to 0.07)	0.8439
Cognitive	0.05 (−0.05 to 0.15)	0.3090	0.04 (−0.09 to 0.17)	0.5450	0.05 (−0.09 to 0.19)	0.5131	−0.01 (−0.11 to 0.06)	0.5903
Neurovegetative	−0.01 (−0.06 to 0.04)	0.6433	−0.01 (−0.07 to 0.06)	0.8245	−0.01 (−0.08 to 0.06)	0.7910	−0.01 (−0.10 to 0.07)	0.6667

GENDEP, Genome-based Therapeutic Drugs for Depression; CI, confidence interval; MADRS, Montgomery–Åsberg Depression Rating Scale; HAMD-17, 17-item Hamilton Depression Rating Scale; BDI, Beck Depression Inventory.

The table shows results for the primary clinician-rated outcome (MADRS), secondary clinician-rated outcome (HAMD-17) and secondary self-report outcome (BDI). Standardized estimates (β), 95% CIs and uncorrected probability of results occurring by chance (p) are based on mixed-effect linear regression models with baseline score on the outcome measure entered as a covariate of no interest (e.g. for all analyses using MADRS as the outcome, MADRS score at baseline was entered as a covariate). β can be interpreted as the effect size. Negative values of β indicate better treatment outcome, positive values of β reflect worse treatment outcome.

p values <0.05 are highlighted in bold to indicate nominal statistical significance.

Table 2. *Prediction of treatment outcome from the six baseline symptom dimensions in GENDEP*

	Overall		Escitalopram		Nortriptyline		Interaction with drug	
	($n=811$)		($n=457$)		($n=354$)		($n=811$)	
Predictor	β (95% CI)	p	β (95% CI)	p	β (95% CI)	p	β (95% CI)	p
(1) Outcome: MADRS								
Mood	0.09 (0.02 to 0.15)	**0.0159**	0.11 (0.02 to 0.19)	**0.0221**	0.05 (−0.06 to 0.15)	0.3785	−0.01 (−0.09 to 0.07)	0.7457
Anxiety	−0.01 (−0.05 to 0.04)	0.6810	−0.06 (−0.11 to 0.01)	0.0874	0.07 (−0.01 to 0.13)	0.0694	0.05 (0.01 to 0.17)	**0.0233**
Pessimism	0.07 (0.02 to 0.12)	**0.0057**	0.07 (0.00 to 0.14)	0.0557	0.07 (−0.01 to 0.13)	0.1024	−0.01 (−0.09 to 0.07)	0.7686
Interest and activity	0.19 (0.13 to 0.24)	$\mathbf{3.9\times10^{-11}}$	0.25 (0.17 to 0.32)	$\mathbf{1.2\times10^{-10}}$	0.13 (0.03 to 0.20)	**0.0061**	−0.04 (−0.15 to 0.01)	0.0813
Sleep	0.05 (0.00 to 0.09)	**0.0477**	0.05 (−0.01 to 0.11)	0.1090	0.05 (−0.02 to 0.12)	0.1591	0.00 (−0.08 to 0.08)	0.9761
Appetite	−0.03 (−0.08 to 0.01)	0.1726	−0.04 (−0.10 to 0.03)	0.2383	−0.03 (−0.10 to 0.04)	0.4339	0.00 (−0.08 to 0.08)	0.9303
(2) Outcome: HAMD-17								
Mood	0.12 (0.06 to 0.16)	$\mathbf{2.9\times10^{-5}}$	0.13 (0.05 to 0.19)	**0.0009**	0.10 (0.01 to 0.17)	**0.0271**	0.00 (−0.09 to 0.07)	0.8724
Anxiety	−0.01 (−0.06 to 0.04)	0.8107	−0.07 (−0.13 to 0.00)	**0.0477**	0.09 (0.01 to 0.16)	**0.0216**	0.06 (0.03 to 0.19)	**0.0073**
Pessimism	0.07 (0.02 to 0.11)	**0.0047**	0.07 (0.00 to 0.13)	**0.0393**	0.06 (−0.02 to 0.12)	0.1610	−0.01 (−0.09 to 0.06)	0.7050
Interest and activity	0.19 (0.14 to 0.23)	$\mathbf{3.3\times10^{-14}}$	0.23 (0.15 to 0.28)	$\mathbf{1.9\times10^{-11}}$	0.15 (0.07 to 0.21)	**0.0002**	−0.03 (−0.13 to 0.02)	0.1413
Sleep	0.04 (−0.01 to 0.09)	0.1085	0.06 (−0.01 to 0.12)	0.0988	0.03 (−0.05 to 0.10)	0.4737	−0.01 (−0.10 to 0.06)	0.6233
Appetite	−0.02 (−0.07 to 0.02)	0.3278	−0.02 (−0.08 to 0.04)	0.5769	−0.04 (−0.11 to 0.03)	0.2878	−0.01 (−0.10 to 0.06)	0.6651
(3) Outcome: BDI								
Mood	0.10 (0.05 to 0.14)	**0.0001**	0.10 (0.03 to 0.16)	**0.0059**	0.11 (0.03 to 0.17)	**0.0056**	0.00 (−0.09 to 0.08)	0.8979
Anxiety	0.04 (0.00 to 0.09)	0.0597	0.04 (−0.03 to 0.09)	0.2604	0.08 (0.00 to 0.14)	**0.0379**	0.01 (−0.06 to 0.11)	0.5772
Pessimism	0.01 (−0.06 to 0.07)	0.7924	0.00 (−0.09 to 0.09)	0.9967	0.01 (−0.09 to 0.10)	0.8790	−0.01 (−0.10 to 0.07)	0.6874
Interest and activity	0.14 (0.07 to 0.19)	$\mathbf{2.1\times10^{-5}}$	0.17 (0.08 to 0.25)	$\mathbf{6.8\times10^{-5}}$	0.11 (0.01 to 0.18)	**0.0248**	−0.03 (−0.13 to 0.04)	0.2579
Sleep	0.07 (0.02 to 0.11)	**0.0035**	0.09 (0.03 to 0.14)	**0.0049**	0.06 (−0.01 to 0.12)	0.0969	−0.02 (−0.13 to 0.04)	0.3077
Appetite	−0.04 (−0.08 to 0.01)	0.1146	−0.04 (−0.10 to 0.02)	0.1922	−0.04 (−0.11 to 0.04)	0.3172	0.00 (−0.09 to 0.08)	0.8654

GENDEP, Genome-based Therapeutic Drugs for Depression; CI, confidence interval; MADRS, Montgomery–Åsberg Depression Rating Scale; HAMD-17, 17-item Hamilton Depression Rating Scale; BDI, Beck Depression Inventory.

The table shows results for the primary clinician-rated outcome (MADRS), secondary clinician-rated outcome (HAMD-17) and secondary self-report outcome (BDI). Standardized estimates (β), 95% CIs and uncorrected probability of results occurring by chance (p) are based on mixed-effect linear regression models with baseline score on the outcome measure entered as a covariate of no interest (e.g. for all analyses using MADRS as the outcome, MADRS score at baseline was entered as a covariate). β is the effect size. Negative values of β indicate better treatment outcome, positive values of β reflect worse treatment outcome.

p values <0.05 are highlighted in bold to indicate nominal statistical significance.

Table 3. *Prediction of treatment outcome from the baseline interest-activity symptom score in STAR*D*

	Model A (primary analysis) (age sex, baseline severity, time in study, clinic)		Model B (with anxiety-somatization) (anxiety-somatization factor score in addition to model A covariates)		Model C (full list of covariates) (ethnicity, marital status, employment, income, age of onset, number of episodes, family history and axis 1 co-morbidity in addition to model A and B covariates)	
Predictor	β (95% CI)	p	β (95% CI)	p	β (95% CI)	p
(1) Outcome: HAMD-17 ($n=2635$)						
Interest-activity	0.09 (0.05–0.14)	**0.0001**	0.09 (0.04–0.14)	**0.0003**	0.09 (0.04 to 0.14)	**0.0002**
Anxiety-somatization	–	–	0.08 (0.03–0.12)	**0.0015**	0.04 (−0.01 to 0.09)	0.0888
(2) Outcome: QIDS-C ($n=3627$)						
Interest-activity	0.36 (0.31–0.42)	$\mathbf{3.5\times10^{-38}}$	0.35 (0.30–0.41)	$\mathbf{5.0\times10^{-36}}$	0.33 (0.28 to 0.39)	$\mathbf{9.8\times10^{-30}}$
Anxiety-somatization	–	–	0.11 (0.06–0.16)	$\mathbf{2.6\times10^{-5}}$	0.05 (−0.01 to 0.10)	0.0832
(3) Outcome: QIDS-SR ($n=3560$)						
Interest-activity	0.24 (0.19–0.29)	$\mathbf{1.5\times10^{-22}}$	0.22 (0.17–0.27)	$\mathbf{2.6\times10^{-19}}$	0.21 (0.16 to 0.26)	$\mathbf{7.7\times10^{-16}}$
Anxiety-somatization	–	–	0.10 (0.06–0.15)	$\mathbf{9.8\times10^{-7}}$	0.05 (0.01 to 0.10)	**0.0209**

STAR*D, Sequenced Treatment Alternatives to Relieve Depression; CI, confidence interval; HAMD-17, 17-item Hamilton Depression Rating Scale; QIDS-C, clinician-rated Quick Inventory of Depression Symptomatology; QIDS-SR, self-report QID.

The table shows results for the primary clinician-rated outcome measure (HAMD-17), secondary clinician-rated outcome measure (QIDS-C) and secondary self-report outcome measure (QIDS-SR). For each outcome, the influence of the interest-activity symptom score is tested in three models with increasing number of covariates (model A, B and C). For models B and C, the results for the previously reported anxiety-somatization symptom score are also presented. Standardized estimates (β), 95% CIs and uncorrected probability of results occurring by chance (p) are based on mixed-effect linear regression models with baseline score on the outcome measure entered as a covariate of no interest (e.g. for all analyses using HAMD-17 as the dependent variable, HAMD-17 score at baseline is entered as a covariate). Negative values of β indicate better treatment outcome, positive values of β reflect worse treatment outcome.

p values <0.05 are highlighted in bold to indicate nominal statistical significance.

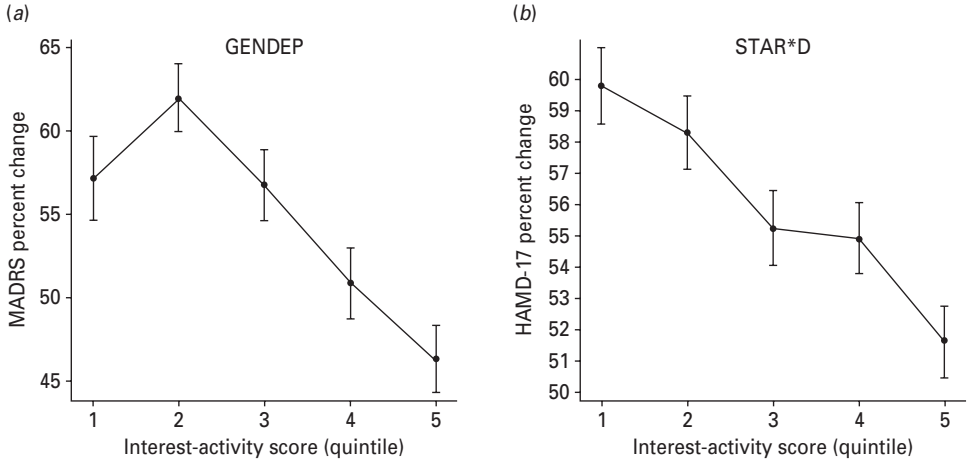

Fig. 2. Association between the interest-activity symptom dimension at baseline and percentage improvement over 12 weeks of treatment on the primary outcome measures in (*a*) Genome-based Therapeutic Drugs for Depression (GENDEP) and (*b*) Sequenced Treatment Alternatives to Relieve Depression (STAR*D). For the purpose of plotting, subjects in each study were separated into quintiles (1 to 5 on the *x* axis) according to increasing interest-activity scores at baseline. The primary outcome measure in GENDEP is the clinician-rated Montgomery–Åsberg Depression Rating Scale (MADRS). The primary outcome measure in STAR*D is the 17-item Hamilton Rating Scale for Depression (HAMD-17) rated by an independent outcome assessor. The percentage reduction on the primary outcome scale over 12 weeks of treatment was adjusted for age, sex and centre differences. Missing week-12 data were imputed by the best unbiased linear estimate from a mixed linear regression model.

The outcome deteriorated gradually with increasing levels of the interest-activity score (Fig. 2). The prediction of outcome by the interest-activity dimension was complementary to the previously reported prediction by the somatization-anxiety score, with both scores independently contributing to the prediction of outcome (Table 3, model B). The strength of the prediction remained unchanged in sensitivity analyses controlling for a comprehensive list of baseline covariates, including ethnicity, marital status, employment, income, age of onset, number of episodes, family history of mood disorder, co-morbid post-traumatic stress and obsessive–compulsive disorder as identified by self-report, number of co-morbid axis I disorders by self-report, and anxiety-somatization score in addition to age, sex, baseline severity and recruiting centre.

Although the prediction was replicated with each of the three outcome measures, there were differences in effect size: the prediction was strongest for the QIDS-C and weakest for HAMD-17. To differentiate effects of scale sensitivity from subject selection, we repeated the analyses with QIDS-C for the 2734 subjects who also had HAMD-17 ratings. We found that the prediction of outcome on QIDS-C in this restricted sample was at least as strong as in the whole sample ($\beta = 0.38$, 95% CI 0.32–0.44, $p = 5.5 \times 10^{-33}$). This excludes selection effect and suggests that the QIDS-C outcome measure may be more sensitive to the prediction of outcome by interest-activity symptoms.

Clinical significance of the prediction

The results reported here establish that the prediction of outcome by the interest-activity symptom dimension is statistically significant and highly unlikely to be due to chance. In addition, we wanted to establish whether the effect size of this prediction was sufficient for applications in clinical settings. For this purpose, we repeated the analyses with the outcome of remission (defined as a HAMD-17 score of $\leqslant 7$ at the last visit in both studies) with no imputation and no covariates. In good agreement with the primary analyses, higher baseline scores on the interest-activity symptoms predicted lower rates of remission in GENDEP [odds ratio (OR) 0.59, 95% CI 0.50–0.68, $p = 1.0 \times 10^{-11}$] and in STAR*D (OR 0.62, 95% CI 0.56–0.67, $p = 3.7 \times 10^{-26}$). Fig. 3 shows that the proportion of individuals who reach remission declines monotonically with increasing baseline scores on the interest-activity symptom dimension. Compared to the lowest scoring fifth of the participants (quintile 1), the rate of remission in the highest scoring one-fifth of participants (quintile 5) was reduced three time in GENDEP and halved in STAR*D (Fig. 3). We next sought to translate this effect into a more clinically useful metric. The AUC was 0.65 in GENDEP and 0.62 in STAR*D (Supplementary Fig. S2), which translates to an NNA of 3 and 4 respectively. In other words, measuring the interest-activity symptom dimension in every three to four patients will help to

https://doi.org/10.1017/S0033291711001905 Published online by Cambridge University Press

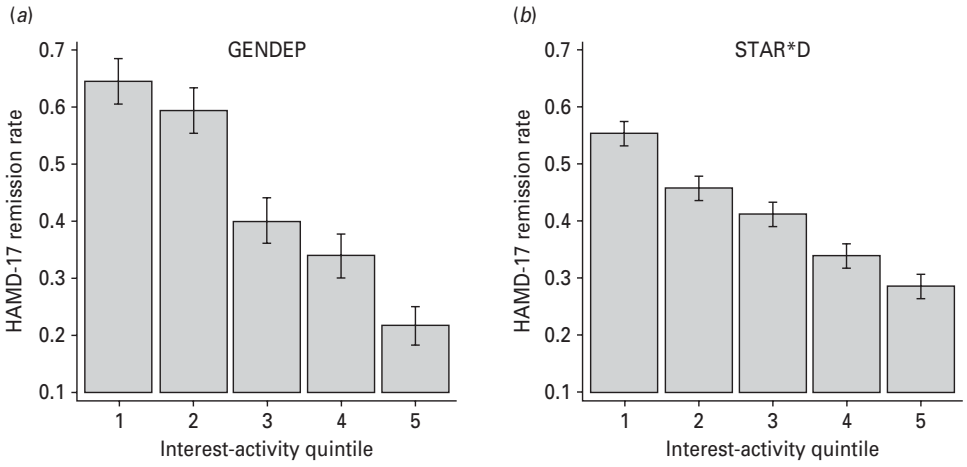

Fig. 3. Association between the interest-activity symptom dimension at baseline and remission [Hamilton Rating Scale for Depression (HAMD)-17 score $\leqslant 7$] in (*a*) Genome-based Therapeutic Drugs for Depression (GENDEP) and (*b*) Sequenced Treatment Alternatives to Relieve Depression (STAR*D). Subjects in each study are separated into quintiles of interest-activity scores at baseline (1–5 on the *x* axis). The proportion reaching remission at last visit is plotted on the *y* axis.

predict one additional remission accurately compared to chance.

Discussion

Investigation of two large treatment trials, one conducted in Europe, the other in the USA, identified robust and consistent evidence that depressed individuals who lack interest, and who are inactive and easily fatigable, experience less improvement during treatment with antidepressant medication, including an SSRI and a noradrenergic tricyclic antidepressant. Comprehensive examination of dimensional predictors in GENDEP found that the interest-activity symptom dimension was the most robust predictor of poor outcome of treatment, irrespective of which antidepressant was used and whether the outcome was measured with a clinician-rated or a self-report scale. This finding was replicated robustly in the STAR*D cohort, where a higher score of interest-activity symptoms at baseline uniquely predicted poor outcome of treatment with citalopram, independently of overall depression severity, age, gender, ethnicity, social class, anxiety and other known predictors of outcome. The effect size of the prediction was substantial, with participants scoring in the highest fifth on the interest-activity symptoms improving by 8–10 percentage points less (Fig. 2) and having, at most, half the chance of achieving remission (Fig. 3) compared those in the lowest fifth in both studies. This finding has implications for clinical care and for research.

Clinical implications

In clinical care, knowledge of outcome predictors may help in forming realistic expectations and in considering alternative approaches for individuals who are less likely to benefit from routine first-line treatment. The results from two large samples representative of subjects treated for depression in routine practice suggest that interest-activity symptoms will help to accurately predict outcome in one additional individual out of every three or four individuals tested.

Given the low burden and cost of measuring depressive symptoms, this predictor can be cost-effective if an alternative treatment is available (Perlis *et al.* 2009). However, clinical application of this finding will require identification of a treatment that is effective in individuals with loss of interest and decreased activity. The interest-activity symptoms at baseline had nearly twice as strong an effect on the outcome of treatment with escitalopram than on the outcome of treatment with nortriptyline in GENDEP. However, the interaction with type of antidepressant was not significant and, for all antidepressants under investigation, diminished interest and activity predicted worse outcome. Therefore, rather than helping choose among different antidepressants, the prominence of interest-activity symptoms may prompt researchers and clinicians to consider alternative or complementary treatment strategies, including alternative pharmacological approaches, behavioural activation or regular exercise. Although no evidence for such treatments in depression with loss of interest and decreased activity is available at present, indirect evidence suggests that several approaches may be effective. For example, behavioural activation directly addresses inactivity, fatigue and lack of involvement. It has been demonstrated that behavioural activation is the effective component of cognitive-behavioural

https://doi.org/10.1017/S0033291711001905 Published online by Cambridge University Press

therapy for depression (Jacobson *et al*. 1996; Dimidjian *et al*. 2006) and that structured psychological therapy is effective in cases where antidepressants have repeatedly failed (Schatzberg *et al*. 2005; Leykin *et al*. 2007). Another complementary modality that may increase activity, improve fatigue and revitalize interest is regular exercise, which has proven efficacy in depression (Mead *et al*. 2009). Given the substantially poorer outcomes in these individuals, they may also merit earlier consideration of pharmacotherapeutic combination or augmentation strategies. Adjunctive treatment with modafinil has been shown to reduce fatigue in cases of depression with low energy that were resistant to antidepressant monotherapy (Rasmussen *et al*. 2005; Thase *et al*. 2006). Future investigations are needed to explore targeted indication and acceptability of these various approaches in depressed individuals, who present with diminished activity, fatigue and loss of interest.

Research implications

The finding that interest-activity symptoms predict response to antidepressants also has implications for future research in depression. New adjunctive treatments may be tested in samples enriched for individuals with high interest-activity symptoms that are less likely to respond to standard treatment. Researchers exploring other predictors and biomarkers of outcome may consider if these are independent of the dimensional structure of symptoms, including interest-activity and anxiety-somatization symptoms (Fava *et al*. 2008). The prediction of outcome may be further improved when such pretreatment predictors are combined with additional variables measured at baseline (Chen *et al*. 2007; Siegle *et al*. 2011) or after a short exposure to antidepressants (Leuchter *et al*. 2009; Bruhl *et al*. 2010). As both early and delayed response to antidepressants occur (Taylor *et al*. 2006; Uher *et al*., in press), symptomatic predictors such as the interest-activity dimension may be even more potent when they are combined with measures of initial improvement during the first weeks of treatment. The identification of activity and fatigue as symptoms relevant to outcome may also suggest biomarkers and molecular mechanisms underlying the individual differences in response. For example, depression with prominent loss of energy and fatigability may be associated with activation of inflammatory pathways that occurs in depression associated with interferon treatment (Raison *et al*. 2009). The interest-activity dimension also constitutes an attractive phenotype for genetic association studies and could enhance pharmacogenetic investigations.

Conceptual implications

More broadly, the results demonstrate the predictive validity of a dimensional classification of depression. In both GENDEP and STAR*D, the outcome deteriorated gradually with increasing levels of the interest-activity score, suggesting that a continuous measure is preferable to a categorical cut-off. This is compatible with the finding that a continuous score of anxiety-somatization was a stronger predictor of outcome than a dichotomy (Fava *et al*. 2008). In addition to interest-activity, cognitive symptoms, mood and anxiety showed a potential for predictive validity in GENDEP. In STAR*D, the interest-activity and anxiety-somatization were complementary and both uniquely contributed to the prediction of outcome. In conjunction with evidence from family (Korszun *et al*. 2004), epidemiological (Bjelland *et al*. 2009; Prisciandaro & Roberts, 2009) and biological investigations (Veen *et al*. 2011; Wardenaar *et al*. 2011), the present results recommend dimensional classification as likely to improve the validity of depression research and support the introduction of dimensional measures in classifications of mental illness.

Strengths and limitations

Several aspects of the methodology deserve comment as they may influence comparability with other reports. First, the symptom dimensions tested in this study were based on rating scales that were designed to assess general depression severity. This means that some aspects of depression that may be important for differentiating depression subtypes, such psychomotor disturbance, were assessed in less detail than what would be possible with specialized scales (Carney *et al*. 1965; Parker, 2007). However, traditional depression subtypes have been the subject of another report (Uher *et al*. 2011) and the present analyses focus on the role of symptom dimensions that are derived empirically rather than based on a particular theoretical background.

Second, both GENDEP and STAR*D used a more inclusive and less tightly controlled design than traditional efficacy trials. This increases the generalizability to routine clinical populations, but introduces the risk of confounding. The flexible dosage of antidepressants means that not every participant receives the same treatment. However, both cognitive and interest-activity symptoms were associated with higher doses of both antidepressants, suggesting that inadequate dosing cannot explain the worse outcome of treatment and that clinicians were adjusting doses upwards in an attempt to achieve satisfactory outcome in the more resistant cases. In addition, the present

finding was robust in a set of sensitivity analyses that controlled for a comprehensive range of potential confounders, suggesting that symptom dimensions uniquely contribute to the outcome of treatment with antidepressants.

Conclusions

In conclusion, convergent findings from two large studies suggest that loss of interest, fatigability, diminished activity and inability to make decisions predict poor outcome of treatment with antidepressants over and above general depression severity and other previously reported predictors of outcome. The substantial effect size of this prediction suggests a potential for clinical application. The prominence of such symptoms may require considering alternative treatment strategies, such as behavioural activation or exercise. Future studies are needed to explore the acceptability and efficacy of such approaches in depression with prominent loss of interest and decreased activity.

Note

Supplementary material accompanies this paper on the Journal's website (http://journals.cambridge.org/psm).

Acknowledgements

The GENDEP project was funded by the European Commission Framework 6 grant, EC Contract Ref.: LSHB-CT-2003-503428. Lundbeck provided nortriptyline and escitalopram for the GENDEP study. GlaxoSmithKline and the UK National Institute for Health Research of the Department of Health contributed to the funding of the sample collection at the Institute of Psychiatry, London. The sponsors had no role in the design and conduct of the study, in data collection, analysis, interpretation or writing the report.

Data for the replication study were obtained from the limited access datasets (version 2) distributed from the NIH-supported STAR*D. STAR*D was supported by National Institute of Mental Health (NIMH) Contract no. N01MH90003 to the University of Texas Southwestern Medical Center. The ClinicalTrials.gov identifier is NCT00021528. This manuscript reflects the views of the authors and may not reflect the opinions or views of the STAR*D Study Investigators or the NIH. R. Uher is supported by a grant from the Innovative Medicines Initiative of the European Commission (Grant Agreement no. 115008). R. H. Perlis is supported by NIMH MH086026.

Declaration of Interest

R. H. Perlis has received consulting fees from Proteus Biomedical, Concordant Rater Systems, and RIDventures. N. Henigsberg has participated in clinical trials sponsored by pharmaceutical companies including GlaxoSmithKline and Lundbeck, and has received honoraria for participating in expert panels from pharmaceutical companies including Lundbeck. D. Souery is a member of the national advisory boards for Astra-Zeneca, Bristol-Myers Squibb, Eli Lilly and Lundbeck. K. J. Aitchison is a member of national advisory boards for Bristol-Myer Squibb and Otsuka Pharmaceuticals Limited and has received speaker's bureau honoraria. The other authors have no conflicts of interests to declare.

References

Beck AT, Ward CH, Mandelson M, Mock J, Erbaugh J (1961). An inventory for measuring depression. *Archives of General Psychiatry* **4**, 561–571.

Bjelland I, Lie SA, Dahl AA, Mykletun A, Stordal E, Kraemer HC (2009). A dimensional versus a categorical approach to diagnosis: anxiety and depression in the HUNT 2 study. *International Journal of Methods in Psychiatric Research* **18**, 128–137.

Brown WA (2007). Treatment response in melancholia. *Acta Psychiatrica Scandinavica* **433**, 125–129.

Bruhl AB, Kaffenberger T, Herwig U (2010). Serotonergic and noradrenergic modulation of emotion processing by single dose antidepressants. *Neuropsychopharmacology* **35**, 521–533.

Carney MWP, Roth M, Garside RF (1965). The diagnosis of depressive syndromes and the prediction of E.C.T. response. *British Journal of Psychiatry* **111**, 659–674.

Chen CH, Ridler K, Suckling J, Williams S, Fu CH, Merlo-Pich E, Bullmore E (2007). Brain imaging correlates of depressive symptom severity and predictors of symptom improvement after antidepressant treatment. *Biological Psychiatry* **62**, 407–414.

Deyi BA, Kosinski AS, Snapinn SM (1998). Power considerations when a continuous outcome variable is dichotomized. *Journal of Biopharmaceutical Statistics* **8**, 337–352.

Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, Gallop R, McGlinchey JB, Markley DK, Gollan JK, Atkins DC, Dunner DL, Jacobson NS (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology* **74**, 658–670.

Fava M, Rush AJ, Alpert JE, Balasubramani GK, Wisniewski SR, Carmin CN, Biggs MM, Zisook S, Leuchter A, Howland R, Warden D, Trivedi MH (2008). Difference in treatment outcome in outpatients with

anxious versus nonanxious depression: a STAR*D report. *American Journal of Psychiatry* **165**, 342–351.

Fava M, Uebelacker LA, Alpert JE, Nierenberg AA, Pava JA, Rosenbaum JF (1997). Major depressive subtypes and treatment response. *Biological Psychiatry* **42**, 568–576.

Flett GL, Vredenburg K, Krames L (1997). The continuity of depression in clinical and nonclinical samples. *Psychological Bulletin* **121**, 395–416.

Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, Rush AJ, Weissman MM (1991). Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. *Archives of General Psychiatry* **48**, 851–855.

Hamilton M (1967). Development of a rating scale for primary depressive illness. *British Journal of Clinical Psychology* **6**, 278–296.

Howland RH, Wilson MG, Kornstein SG, Clayton AH, Trivedi MH, Wohlreich MM, Fava M (2008). Factors predicting reduced antidepressant response: experience with the SNRI duloxetine in patients with major depression. *Annals of Clinical Psychiatry* **20**, 209–218.

Jacobson NS, Dobson KS, Truax PA, Addis ME, Koerner K, Gollan JK, Gortner E, Prince SE (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology* **64**, 295–304.

Joyce PR, Mulder RT, McKenzie JM, Luty SE, Cloninger CR (2004). Atypical depression, atypical temperament and a differential antidepressant response to fluoxetine and nortriptyline. *Depression and Anxiety* **19**, 180–186.

Korszun A, Moskvina V, Brewster S, Craddock N, Ferrero F, Gill M, Jones IR, Jones LA, Maier W, Mors O, Owen MJ, Preisig M, Reich T, Rietschel M, Farmer A, McGuffin P (2004). Familiality of symptom dimensions in depression. *Archives of General Psychiatry* **61**, 468–474.

Kraemer HC, Kupfer DJ (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry* **59**, 990–996.

Leuchter AF, Cook IA, Gilmer WS, Marangell LB, Burgoyne KS, Howland RH, Trivedi MH, Zisook S, Jain R, Fava M, Iosifescu D, Greenwald S (2009). Effectiveness of a quantitative electroencephalographic biomarker for predicting differential response or remission with escitalopram and bupropion in major depressive disorder. *Psychiatry Research* **169**, 132–138.

Leykin Y, Amsterdam JD, DeRubeis RJ, Gallop R, Shelton RC, Hollon SD (2007). Progressive resistance to a selective serotonin reuptake inhibitor but not to cognitive therapy in the treatment of major depression. *Journal of Consulting and Clinical Psychology* **75**, 267–276.

Mallinckrodt CH, Zhang L, Prucka WR, Millen BA (2010). Signal detection and placebo response in schizophrenia: parallels with depression. *Psychopharmacology Bulletin* **43**, 53–72.

McGrath PJ, Khan AY, Trivedi MH, Stewart JW, Morris DW, Wisniewski SR, Miyahara S, Nierenberg AA, Fava M, Rush JA (2008). Response to a selective serotonin reuptake inhibitor (citalopram) in major depressive disorder with melancholic features: a STAR*D report. *Journal of Clinical Psychiatry* **69**, 1847–1855.

McGrath PJ, Stewart JW, Janal MN, Petkova E, Quitkin FM, Klein DF (2000). A placebo-controlled study of fluoxetine versus imipramine in the acute treatment of atypical depression. *American Journal of Psychiatry* **157**, 344–350.

Mead GE, Morley W, Campbell P, Greig CA, McMurdo M, Lawlor DA (2009). Exercise for depression. *Cochrane Database of Systematic Reviews* (4), CD004366.

Montgomery SA, Åsberg M (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry* **134**, 382–389.

Nelson JC (2010). Anxiety does not predict response to duloxetine in major depression: results of a pooled analysis of individual patient data from 11 placebo-controlled trials. *Depression and Anxiety* **27**, 12–18.

Parker G (2007). Defining melancholia: the primacy of psychomotor disturbance. *Acta Psychiatrica Scandinavica. Supplementum* (433), 21–30.

Parker G, Fletcher K, Hyett M, Hadzi-Pavlovic D, Barrett M, Synnott H (2009). Measuring melancholia: the utility of a prototypic symptom approach. *Psychological Medicine* **39**, 989–998.

Parker G, Wilhelm K, Mitchell P, Roy K, Hadzi-Pavlovic D (1999). Subtyping depression: testing algorithms and identification of a tiered model. *Journal of Nervous and Mental Disease* **187**, 610–617.

Paykel ES, Rowan PR, Parker RR, Bhat AV (1982). Response to phenelzine and amitriptyline in subtypes of outpatient depression. *Archives of General Psychiatry* **39**, 1041–1049.

Perlis RH, Patrick A, Smoller JW, Wang PS (2009). When is pharmacogenetic testing for antidepressant response ready for the clinic? A cost-effectiveness analysis based on data from the STAR*D study. *Neuropsychopharmacology* **34**, 2227–2236.

Perry PJ (1996). Pharmacotherapy for major depression with melancholic features: relative efficacy of tricyclic versus selective serotonin reuptake inhibitor antidepressants. *Journal of Affective Disorders* **39**, 1–6.

Prisciandaro JJ, Roberts JE (2005). A taxometric investigation of unipolar depression in the National Comorbidity Survey. *Journal of Abnormal Psychology* **114**, 718–728.

Prisciandaro JJ, Roberts JE (2009). A comparison of the predictive abilities of dimensional and categorical models of unipolar depression in the National Comorbidity Survey. *Psychological Medicine* **39**, 1087–1096.

Raison CL, Borisov AS, Majer M, Drake DF, Pagnoni G, Woolwine BJ, Vogt GJ, Massung B, Miller AH (2009). Activation of central nervous system inflammatory pathways by interferon-alpha: relationship to monoamines and depression. *Biological Psychiatry* **65**, 296–303.

Rasmussen NA, Schroder P, Olsen LR, Brodsgaard M, Unden M, Bech P (2005). Modafinil augmentation in depressed patients with partial response to antidepressants: a pilot study on self-reported symptoms covered by the Major Depression Inventory (MDI) and the Symptom Checklist (SCL-92). *Nordic Journal of Psychiatry* **59**, 173–178.

Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH (1996). The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychological Medicine* **26**, 477–486.

Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH, Keller MB (2003). The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry* **54**, 573–583.

Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackeim HA, Kupfer DJ, Luther J, Fava M (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *American Journal of Psychiatry* **163**, 1905–1917.

Russell JM, Koran LM, Rush J, Hirschfeld RM, Harrison W, Friedman ES, Davis S, Keller M (2001). Effect of concurrent anxiety on response to sertraline and imipramine in patients with chronic depression. *Depression and Anxiety* **13**, 18–27.

Schatzberg AF, Rush AJ, Arnow BA, Banks PL, Blalock JA, Borian FE, Howland R, Klein DN, Kocsis JH, Kornstein SG, Manber R, Markowitz JC, Miller I, Ninan PT, Rothbaum BO, Thase ME, Trivedi MH, Keller MB (2005). Chronic depression: medication (nefazodone) or psychotherapy (CBASP) is effective when the other is not. *Archives of General Psychiatry* **62**, 513–520.

Siegle GJ, Steinhauer SR, Friedman ES, Thompson WS, Thase ME (2011). Remission prognosis for cognitive therapy for recurrent depression using the pupil: utility and neural correlates. *Biological Psychiatry* **69**, 726–733.

Stewart JW, McGrath PJ, Fava M, Wisniewski SR, Zisook S, Cook I, Nierenberg AA, Trivedi MH, Balasubramani GK, Warden D, Lesser I, John RA (2010). Do atypical features affect outcome in depressed outpatients treated with citalopram? *International Journal of Neuropsychopharmacology* **13**, 15–30.

Streiner DL (2002). Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Canadian Journal of Psychiatry* **47**, 262–266.

Taylor MJ, Freemantle N, Geddes JR, Bhagwagar Z (2006). Early onset of selective serotonin reuptake inhibitor antidepressant action: systematic review and meta-analysis. *Archives of General Psychiatry* **63**, 1217–1223.

Thase ME (2009). Atypical depression: useful concept, but it's time to revise the DSM-IV criteria. *Neuropsychopharmacology* **34**, 2633–2641.

Thase ME, Fava M, DeBattista C, Arora S, Hughes RJ (2006). Modafinil augmentation of SSRI therapy in patients with major depressive disorder and excessive sleepiness and fatigue: a 12-week, open-label, extension study. *CNS Spectrums* **11**, 93–102.

Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, Norquist G, Howland RH, Lebowitz B, McGrath PJ, Shores-Wilson K, Biggs MM, Balasubramani GK, Fava M (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *American Journal of Psychiatry* **163**, 28–40.

Uher R, Dernovsek MZ, Mors O, Hauser J, Souery D, Zobel A, Maier W, Henigsberg N, Kalember P, Rietschel M, Placentino A, Mendlewicz J, Aitchison KJ, McGuffin P, Farmer A (2011). Melancholic, atypical and anxious depression subtypes and outcome of treatment with escitalopram and nortriptyline. *Journal of Affective Disorders* **132**, 112–120.

Uher R, Farmer A, Maier W, Rietschel M, Hauser J, Marusic A, Mors O, Elkin A, Williamson RJ, Schmael C, Henigsberg N, Perez J, Mendlewicz J, Janzing JG, Zobel A, Skibinska M, Kozel D, Stamp AS, Bajs M, Placentino A, Barreto M, McGuffin P, Aitchison KJ (2008). Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychological Medicine* **38**, 289–300.

Uher R, Maier W, Hauser J, Marusic A, Schmael C, Mors O, Henigsberg N, Souery D, Placentino A, Rietschel M, Zobel A, Dmitrzak-Weglarz M, Petrovic A, Jorgensen L, Kalember P, Giovannini C, Barreto M, Elkin A, Landau S, Farmer A, Aitchison KJ, McGuffin P (2009*a*). Differential efficacy of escitalopram and nortriptyline on dimensional measures of depression. *British Journal of Psychiatry* **194**, 252–259.

Uher R, Mors O, Hauser J, Rietschel M, Maier W, Kozel D, Henigsberg N, Souery D, Placentino A, Perroud N, Dernovsek MZ, Strohmaier J, Larsen ER, Zobel A, Leszczynska-Rodziewicz A, Kalember P, Pedrini L, Linotte S, Gunasinghe C, Aitchison KJ, McGuffin P, Farmer A (2009*b*). Body weight as a predictor of antidepressant efficacy in the GENDEP project. *Journal of Affective Disorders* **118**, 147–154.

Uher R, Mors O, Rietschel M, Rajewska-Rager A, Petrovic A, Zobel A, Henigsberg N, Mendlewicz J, Aitchison KJ, Farmer A, McGuffin P (in press). Early and delayed onset of response to antidepressants in individual trajectories of change during treatment of major depression. *Journal of Clinical Psychiatry*.

Uher R, Muthen B, Souery D, Mors O, Jaracz J, Placentino A, Petrovic A, Zobel A, Henigsberg N, Rietschel M, Aitchison KJ, Farmer A, McGuffin P (2010). Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine* **40**, 1367–1377.

Veen G, van Vliet IN, Derijk RH, Giltay EJ, van Pelt J, Zitman FG (2011). Basal cortisol levels in relation to dimensions and DSM-IV categories of depression and anxiety. *Psychiatry Research* **185**, 121–128.

Wardenaar KJ, Vreeburg SA, van Veen T, Giltay EJ, Veen G, Penninx BW, Zitman FG (2011). Dimensions of depression and anxiety and the hypothalamo-pituitary-adrenal axis. *Biological Psychiatry* **69**, 366–373.

Wing JK, Sartorius N, Ustin TB (1998). *Diagnosis and Clinical Measurement in Psychiatry. A Reference Manual for SCAN*. World Health Organization: Geneva.